



# Interaction matrix selection in spatial econometrics with an application to growth theory

Nicolas Debarsy, Cem Ertur

## ► To cite this version:

Nicolas Debarsy, Cem Ertur. Interaction matrix selection in spatial econometrics with an application to growth theory. *Regional Science and Urban Economics*, 2019, 75, pp.49-69. <10.1016/j.regsciurbeco.2019.01.002>. <halshs-01278545v3>

**HAL Id: halshs-01278545**

**<https://shs.hal.science/halshs-01278545v3>**

Submitted on 30 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Interaction matrix selection in spatial autoregressive models with an application to growth theory

Nicolas Debarsy<sup>a,1,\*</sup>, Cem Ertur<sup>2</sup>

<sup>a</sup>*CNRS, LEO CNRS FRE 2014, Rue de Blois, BP 26739, F45067, Orléans*

---

## Abstract

The interaction matrix, or spatial weight matrix, is the fundamental tool to model cross-sectional interdependence between observations in spatial autoregressive models. However, it is most of the time not derived from theory, as it should be ideally, but chosen on an ad hoc basis. In this paper, we propose a modified version of the  $J$  test to formally select the interaction matrix. Our methodology is based on the application of the robust against unknown heteroskedasticity GMM estimation method, developed by Lin and Lee (2010). We then implement the testing procedure developed by Hagemann (2012) to overcome the decision problem inherent to non-nested models tests.

An application of the testing procedure is presented for the Schumpeterian growth model with world-wide interactions developed by Ertur and Koch (2011) using three different types of interaction matrices: genealogic distance, linguistic distance and bilateral trade flows. We find that the interaction matrix based on trade flows is the most adequate.

*Keywords:* Bootstrap, GMM, Interaction matrix,  $J$  tests, Non-nested models, Heteroskedasticity, SAR  
*JEL:* : C21, C12, C52, O4

---

## 1. Introduction

The interaction matrix, or spatial weight matrix, is the fundamental tool needed to model cross-sectional interdependence between observations in spatial econometric models. It allows to specify the structure of interaction or connectivity within a system through a spatial econometric model and is presumed to be known and exogenous (Anselin and Bera, 1998). However, it is most of the time not derived from theory, as it should be ideally, but chosen on an ad hoc basis. In practice, an applied researcher has to choose a priori the interaction matrix, most of the time without any theoretical guidance. As a result, in empirical studies, interaction matrices are generally based on geographical distance, whose exogeneity is widely admitted.

---

\*Corresponding author

*Email address:* nicolas.debarsy@cnrs.fr (Nicolas Debarsy)

<sup>1</sup>Permanent Address: CNRS, LEM- UMR 9221, Université de Lille, Cité Scientifique, Bat. SH2 F-59655 Villeneuve-d'Ascq

<sup>2</sup>The current version of this paper was completed after Cem Ertur's death, on October 9 2016.. I want to express my deep sorrow for the loss of a close friend and an excellent economist, from whom I had learned much over the years.

Nevertheless, other concepts of similarity or proximity can be useful to explain interactions. Indeed, by analogy to Akerlof (1997), countries or regions can be seen as localized in some general socio-economic, institutional or political space defined by a set of factors. Implementation of spatial econometric models thus requires the proper definition of the similarity or proximity measure to be used for building the interaction matrix.

Moreover, the consequences of a misspecification of the interaction matrix can be severe. For example, Lee (2008) shows that the misspecification of the interaction matrix can cause a significant bias in the estimates of the spatial autoregressive parameter in a SAR model. If the interaction matrix is under-specified, then both the two stage least squares (2SLS) and maximum likelihood (ML) estimator of the spatial autoregressive parameter are biased downward. Inversely, if the interaction matrix is over-specified, the spatial autoregressive parameter is overestimated. Furthermore, the under-specification of the interaction matrix is more problematic than its over-specification in terms of both bias and RMSE.

A special attention should therefore be given to the choice of the interaction matrix since it conditions the estimated values of the parameters as well as the impacts (and/or inference) of spatial autoregressive models. One of the key points in the specification of spatial econometric models is then the choice, given a criterion, of the most relevant interaction matrix in a subset of all possible matrices.

As different types of interaction structures can be specified to study the same phenomenon, it can prove difficult to identify the most appropriate interaction matrix. The literature on the tests of non-nested models has nevertheless made possible the handling of such situations. The literature devoted to the tests of non-nested assumptions finds its origins in many papers including those of Cox (1961, 1962); Atkinson (1970); Davidson and MacKinnon (1981); Pesaran (1974); Pesaran and Deaton (1978); MacKinnon et al. (1983); Godfrey (1983) and Pesaran and Weeks (2001). The tests of non-nested models are classified in two broad categories. The first one is the test of Cox (1961, 1962), adapted to linear regression models by Pesaran (1974) and to non-linear regression models by Pesaran and Deaton (1978). The second category is due to Davidson and MacKinnon (1981). These authors develop several procedures, closely related to each others, which may be used to test against several alternative models simultaneously.

In the spatial econometrics literature, one can find a first contribution in Anselin (1988). More recently,

Kelejian (2008) extended the  $J$  test of Davidson and MacKinnon (1981) to a spatial context.<sup>3</sup> The principle of the  $J$  test is to estimate a model which contains the null model and the predictive power of alternative specifications. Then the significance of additional terms is tested. Since this test does not use available information in an efficient way, Kelejian and Piras (2011) and Burridge (2012) suggest different improvements. Kelejian and Piras (2011) modify the way available information is used in determining the predictions and consequently improve the power of the  $J$  test. Burridge (2012) advocates the use of parameter estimates obtained from maximum likelihood estimation. Finally in an alternative approach, Jin and Lee (2013) extend the Cox test (Cox, 1961, 1962) to a spatial framework. Moreover, in the closely related econometric network literature, an extension of the Kelejian’s (2008)  $J$  test is proposed by Liu et al. (2014) to differentiate between the local-aggregate and the local-average endogenous peer effects in an econometric network model with network fixed effects.

In this paper, we propose a modified version of the  $J$  test based on the robust to unknown heteroskedasticity Generalized Method of Moments (RGMM) estimation method derived by Lin and Lee (2010). It is well known in the literature that all the non-nested hypotheses tests suffer from a decision problem: in some circumstances, they may not help discriminating between the null and alternative models. To overcome this decision problem, Hagemann (2012) recently suggests a simple test for non-nested models, namely the  $MJ$  test, that we present here in the framework of spatial econometric models.

In this work, we focus on selecting the most adequate connectivity matrix among a set of candidates. However the literature also developed estimation procedures to account for several connectivity matrices within a spatial autoregressive specification. Firstly, Lacombe (2004); Badinger and Egger (2011); Lee and Liu (2010); Elhorst et al. (2012) and Liu et al. (2014) among others have proposed spatial autoregressive models introducing multiple spatial lags of the dependent variable as right-hand side variables to create a ‘higher-order’ SAR specification. Secondly, Pace and LeSage (2002); Hazır et al. (2018) and Debarsy and LeSage (2018) among others have developed alternative spatial autoregressive specifications that relies on convex combinations of different connectivity matrices to form a single weight matrix. However these two types of generalizations suffer from some limits. Regarding the higher-order SAR model, there is no consensus in the literature regarding the parameter space to be used when estimating several spatial autoregressive parameters. Elhorst et al. (2012) show that the parameter space proposed by Lee and

---

<sup>3</sup>Not to be confused with the Hansen-Sargan  $J$  test for over-identification in GMM estimation.

Liu (2010) and Badinger and Egger (2011) is not wide enough and, quoting Sherell (1990), indicate that the parameter space is difficult to find. With respect to the convex combination approach, the estimation procedures have been developed under the maximum likelihood framework, which imposes homoskedasticity of the error term. By contrast, our proposed approach, based on one single interaction matrix, has a well-defined parameter space for the spatial autoregressive parameter and allows for unknown heteroskedasticity.

We finally apply the preceding non-nested hypothesis tests to the generalized Schumpeterian growth model proposed by Ertur and Koch (2011). These authors propose a theoretical and methodological framework taking into account technological interactions between countries to explain the growth process in a Schumpeterian perspective. Worldwide interdependence, implied by international R&D spillovers, are therefore structurally introduced in a growth model with technological interactions, which yields a spatial econometric reduced form. This model includes both factor accumulation and innovation caused by R&D investments as determinants of the growth process, with technological diffusion occurring between interacting pairs of countries, human capital, reflecting the capacity for absorption of technology and physical capital playing the usual part.

In this paper, we consider three different interaction schemes to model the links between countries. The first one is based on genealogical or genetic distance, also called coancestry coefficient (Reynolds et al., 1983). This concept of distance, defined by Spolaore and Wacziarg (2009, 2013), based on Cavalli-Sforza et al. (1994), corresponds to the time elapsed since two populations shared a common ancestor.<sup>4</sup>

In this setting, genealogical distance is considered as a proxy for cultural and institutional similarities. The results of Spolaore and Wacziarg (2009) suggest that this general measure of genealogical relatedness between populations can explain cross-country income differences. Their interpretation is that genealogical distance captures obstacles to the diffusion of development. The second interaction matrix we consider is based on linguistic proximity, following Melitz and Toubal (2014). Melitz and Toubal (2014) summarize the evidence about the linguistic influences in an index, named Common Language Index (CL), resting strictly on exogenous linguistic factors. This summary index (CL) is constructed from different proxies: Common Official Language (COL), Common Native Language (CNL) and Language Proximity (LP). The third

---

<sup>4</sup>For other empirical works taking into account the ancestral composition of current populations see also Putterman and Weil (2010); Comin et al. (2010); Ashraf and Galor (2013).

interaction matrix used is based on trade flows, which may proxy multi-country technological interactions.

The use of the  $MJ$  test provides considerable help in the selection of the interaction matrix. Indeed, in the context of this application, we show that the trade flows based interaction matrix is the most appropriate pattern to model interactions between countries, among the three types of interaction patterns we consider.

The remainder of the paper is organized as follows: in section 2, we briefly discuss the estimation procedure, namely the RGMM approach developed by Lin and Lee (2010) and present model interpretation. Our modified version of the  $J$  test based on RGMM estimation as well as the predictor selection are then presented in section 3. The  $MJ$  test framework developed by Hagemann (2012) is then applied to our specific context of interaction matrix selection in section 4. In section 5, we show the good performance of our proposed tests in small samples through Monte Carlo experiments. Section 6 presents the application to the generalized Schumpeterian growth model while section 7 concludes.

## 2. Heteroskedasticity-robust estimation of SAR model

The model considered in this paper is the heteroskedastic SAR specification

$$Y_n = X_n\beta_0 + \lambda_0 W_n Y_n + \varepsilon_n. \quad (1)$$

where  $X_n$  is a  $n \times k$  matrix of exogenous and non-stochastic explanatory variables,  $W_n$  is a non-stochastic interaction matrix whose diagonal elements are zero,  $\lambda_0$  is the true parameter measuring interactions' intensity, labeled spatial autoregressive or cross-sectional dependence parameter, while  $\beta_0$  is the true value of the  $k \times 1$  vector of unknown parameters associated to the explanatory variables. Finally,  $\varepsilon_n$  is the vector of independent error terms with  $\varepsilon_{in}$  characterized by a zero mean and a variance  $\sigma_{in}^2$ ,  $i = 1, \dots, n$ . We also note  $\theta_0 = [\lambda_0, \beta_0']'$  and  $\theta = [\lambda, \beta']'$  for any value of  $\lambda$  and  $\beta$ .<sup>5</sup>

Lin and Lee (2010) show that the (quasi-) maximum likelihood estimator (QMLE) for  $\lambda$  of model (1) is generally inconsistent in presence of unknown heteroskedasticity. This inconsistency comes from the definition of the interaction scheme used. Without entering the technical details (which can be found in

---

<sup>5</sup>In this paper, we note  $\lambda_0$  and  $\beta_0$  as the true value of the parameters,  $\hat{\lambda}$  and  $\hat{\beta}$  as estimators of  $\lambda_0$  and  $\beta_0$  and  $\lambda$  and  $\beta$  as parameters on which we maximize the objective function. Hence, only  $\hat{\lambda}$  and  $\hat{\beta}$  are viewed as random variables.

Lin and Lee (2010) and Liu and Yang (2015)), one needs the covariance between diagonal elements of the matrix  $G_n$ ,  $G_{n,ii}, i = 1, \dots, n$  and individual variances  $\sigma_{in}^2, i = 1, \dots, n$  to be zero in the limit to have a consistent QMLE for  $\lambda$ , where  $G_n = W_n(I_n - \lambda_0 W_n)^{-1}$ . When imposing homoskedasticity of the error term, this covariance is by definition 0 and the QMLE of  $\lambda$  is consistent. When heteroskedasticity of errors is allowed, one needs (most) of the diagonal elements of  $G_n$  to be constant (in the limit) across observations. However, these elements are functions of  $W_n$ , meaning that depending on the structure of the interaction scheme, a consistent QMLE for  $\lambda$  might be obtained. Lin and Lee (2010) and Liu and Yang (2015) show the conditions under which consistency is achieved and further present empirical cases: the group interaction structure with groups of the same size or the “circular” world where the units are arranged on a circle such that the last unit  $y_n$  has neighbors  $y_1$  and  $y_{n-1}$ ,  $y_1$  has neighbors  $y_2$  and  $y_n$ . However, for general  $W_n$  matrices, the QMLE is inconsistent. It is important to note that the inconsistency of QMLE in presence of unknown heteroskedasticity comes from the structure of  $W_n$ . If the diagonal elements of  $G_n$  are asymptotically constant, the QMLE of the SAR model derived under homoskedasticity can still be consistent with heteroskedastic errors. Based on this result, Liu and Yang (2015) derive a modification of the concentrated score function for the spatial parameter to make it robust against unknown heteroskedasticity. They further propose an outer-product-of-gradient method for estimating the variance of the modified QMLE and obtain heteroskedasticity robust inference.

Alternative estimation methods like the Two-Stage Least Squares (TSLS) approach (Kelejian and Prucha, 1998), the Best TSLS (Lee, 2003) and the series Best TSLS (Kelejian et al., 2004) could be considered, but according to Kelejian and Prucha (2010, p.60), the asymptotic normality results obtained for TSLS under homoskedasticity does not carry over in presence of heteroskedasticity. Kelejian and Prucha (2010) generalize the method introduced in Kelejian and Prucha (1999) to estimate a SAR model with SAR disturbances in the presence of unknown heteroskedasticity. They combine a generalized method of moments estimator (GMME) for the coefficient of spatial autocorrelation in the error term to a TSLS approach for the coefficient associated to the spatial lag of the dependent variable.

In this paper, we rely on the GMM estimation procedure developed by Lin and Lee (2010) to estimate (1).<sup>6</sup> The advantage of GMM over TSLS is that the former remains consistent even if none of the

---

<sup>6</sup>We might also have used the modified QML estimation procedure of Liu and Yang (2015) but the GMM framework further allows to easily consider additional endogenous explanatory variables in the specification.

explanatory variable is relevant (Lee, 2007). Besides TSLS only uses the information contained in the deterministic part of the reduced form of (1), leaving the information in the stochastic part of the reduced form unexploited. Hence, the GMME will be more efficient than the TSLSE. Lin and Lee (2010) generalize Lee (2007) and Liu et al. (2010) developed for homoskedastic errors. The main differences in the estimation procedure when accounting for heteroskedastic errors come from the constraint on quadratic moments and the absence of a Best Optimal GMME.<sup>7</sup>

Let us now consider the interpretation of the estimation results of the heteroskedastic SAR model in terms of the impact of a variation of an independent variable on the dependent variable. As the model (1) is estimated in implicit form, we need to rely on its reduced form, shown in (2), to provide economic interpretations.

$$Y_n = \sum_{h=1}^k \beta_{0,h} (I_n - \lambda_0 W_n)^{-1} X_{n,h} + (I_n - \lambda_0 W_n)^{-1} \varepsilon_n \quad (2)$$

Taking the first order partial derivatives of  $Y_n$  with respect to  $X_{n,h}$ , for  $h = 1, \dots, k$ , we obtain  $k$  ( $n \times n$ ) impact matrices, one for each of the explanatory variables:

$$\Xi_n^{(h)} = \frac{\partial Y_n}{\partial X_{n,h}'} = \beta_{0,h} (I_n - \lambda_0 W_n)^{-1} = \beta_{0,h} \sum_{v=0}^{\infty} \lambda_0^v W_n^v = \beta_{0,h} (I_n + \lambda_0 W_n + \lambda_0^2 W_n^2 + \lambda_0^3 W_n^3 + \dots) \quad (3)$$

where  $(I_n - \lambda_0 W_n)^{-1}$  is the so-called *global interaction multiplier*. Note that the impact matrices are generally full and not symmetric regardless of the sparsity and structure of the interaction matrix  $W_n$ . We call the country in column  $j$  of this matrix the emitting country and country in row  $i$  the receiving country.

The diagonal elements of the impact matrix, specified in equation (3), called the direct impacts, are heterogeneous in presence of spatial autocorrelation due to higher order feedback effects. More precisely, the own derivative for country  $i$  includes feedback effects where country  $i$  affects country  $j$  and country  $j$  also affects country  $i$  as well as longer paths which might go from country  $i$  to  $j$  to  $k$  and back to  $i$ . This is what Debarsy and Ertur (2010) call *interactive heterogeneity*, by contrast to standard individual

---

<sup>7</sup>We provided a detailed implementation of the Lin and Lee (2010) RGMM estimator in Appendix A.



heterogeneity in panel data models.

$$Diag(\Xi_n^{(h)}) = \beta_{0,h}I_n + \beta_{0,h}Diag(\lambda_0^2W_n^2 + \lambda_0^3W_n^3 + \dots) \quad (4)$$

The magnitude of those direct effects mostly depends on the value of  $\beta_{0,h}$ , constant across the sample. Heterogeneity thus comes from the second term on the right hand side of (4) and represents the magnitude of pure feedback effects, which depends again on  $\beta_{0,h}$ , on the values of powers of  $\lambda_0$ , and on the interconnection structure of the observations embedded in the powers of the interaction matrix  $W_n$ . This heterogeneity is nevertheless likely to be negligible compared to the value of  $\beta_{0,h}$  in most of applied works.

The main question spatial autoregressive models allow to answer concerns the impact of a variation of an explanatory variable in a country  $i$  on the dependent variable in other countries of the sample, which are called indirect effects. They correspond to the off-diagonal terms of the impact matrix, and represent the *spillovers*. Their expression is shown in (5):

$$\Xi_n^{(h)} - Diag(\Xi_n^{(h)}) = \beta_{0,h}\lambda_0W_n + \beta_{0,h}[\lambda_0^2W_n^2 - Diag(\lambda_0^2W_n^2)] + \beta_{0,h}[\lambda_0^3W_n^3 - Diag(\lambda_0^3W_n^3)] + \dots \quad (5)$$

Finally, the sum of the  $i^{th}$  row of the impact matrix (3) represents the total impact on the dependent variable in country  $i$  due to a 1 unit change in  $X_{n,h}$  in all of the countries in the sample. The sum of column  $j$  gives the total impact on the dependent variable of all the countries of a 1 unit change of  $X_{n,h}$  in country  $j$ , which is of particular interest in terms of interpretation here.

### 3. The modified $J$ test based on RGMM

The general motivation of the  $J$  test is to estimate a model which contains the model under the null hypothesis (considered as the correct model) and the predictive power of the alternative models, and assess if predictors from these alternative specifications can significantly contribute to the explanatory power in the null model. For ease of exposition, the model under the null hypothesis  $H_0$  and the  $M - 1$  alternative hypotheses  $H_{1,m}, m = 2, \dots, M$ , are heteroskedastic SAR specifications which only differ by the interaction

matrix used:<sup>8</sup>

$$H_0 : Y_n = X_n\beta_1 + \lambda_1 W_{1n}Y_n + \varepsilon_{1n}, \quad (6)$$

$$H_{1,m} : Y_n = X_n\beta_m + \lambda_m W_{mn}Y_n + \varepsilon_{mn}, m = 2, \dots, M \quad (7)$$

For the homoskedastic case, different predictors have been derived by Kelejian and Prucha (2007), presented here for the sake of comparison with the heteroskedastic case.

### 3.1. Predictors for homoskedastic errors

The first predictor for model  $m$   $\hat{y}_{in|1}^m$  is obtained from its reduced form:

$$\begin{aligned} \hat{y}_{in|1}^m &= \mathbf{E}(y_{in}|\Lambda_1) \\ &= (I_n - \lambda_m W_{mn})_{i.}^{-1} X_n \beta_m, \end{aligned} \quad (8)$$

where  $A_{i.}$  means the  $i^{th}$  row of matrix  $A$ . This predictor is based on the smallest information set, namely  $\Lambda_1 = \{X_n, W_{mn}\}$ . The second predictor is based on a larger information set, namely  $\Lambda_2 = \{X_n, W_{mn}, W_{mn,i}Y_n\}$  and is thus more efficient. It is obtained from the structural form and is computed as follows:

$$\begin{aligned} \hat{y}_{in|2}^m &= \mathbf{E}(y_{in}|\Lambda_2) \\ &= \lambda_m W_{mn,i}Y_n + X_{i.,n}\beta_m + \frac{cov(\varepsilon_{in}, W_{mn,i}Y_n)}{var(W_{mn,i}Y_n)} [W_{mn,i}Y_n - \mathbf{E}(W_{mn,i}Y_n)] \end{aligned} \quad (9)$$

The last term of (9) accounts for the correlation between the error term and the spatial lag of the dependent variable. The different elements involved in this correlation are computed as follows:

$$\begin{aligned} \mathbf{E}(W_{mn,i}Y_n) &= W_{mn,i}(I - \lambda_m W_{mn})^{-1} X_n \beta_m \\ \Sigma_m^y &= (I - \lambda_m W_{mn})^{-1} (I - \lambda_m W_{mn}')^{-1} \\ var(W_{mn,i}Y_n) &= \sigma_n^2 W_{mn,i} \Sigma_m^y W_{mn,i}' \\ cov(\varepsilon_{in}, W_{mn,i}Y_n) &= \sigma_n^2 (I - \lambda_m W_{mn}')^{-1} X_n \beta_m \end{aligned} \quad (10)$$

---

<sup>8</sup>The  $J$  test approach proposed here can also be used to additionally test for the model specification. However, this would lead to a joint hypothesis problem. As such, when the  $J$  test is rejected, one cannot conclude whether this is due to the interaction scheme used, the different model specification considered or both.

where  $\sigma_n^2$  is the variance of the error term. A third predictor is sometimes considered due to its intuitive appeal. It is also based on the information set  $\Lambda_2$  but assumes a zero correlation between the error term and the spatial lag of the dependent variable:

$$\hat{y}_{in|3}^m = \lambda_m W_{mn,i} Y_n + X_{i.,n} \beta_m.$$

Kelejian and Prucha (2007) show that it is biased since the spatial lag is assumed independent of the error term. We thus do not consider it.<sup>9</sup>

### 3.2. Predictors for heteroskedastic errors

The first predictor,  $\hat{y}_{in|1}^m$ , is not affected by the presence of heteroskedasticity and takes the same expression as for homoskedastic errors (see 8). However, in the second predictor,  $\hat{y}_{in|2}^m$ , the variance of the spatial lag and the covariance between the spatial lag and the error term are affected by heteroskedasticity, as shown in (11):

$$\begin{aligned} \text{var}(W_{mn,i} Y_n) &= \sigma_{i,mn}^2 W_{m,i} \Sigma_m^y W'_{mn,i}, \\ \text{cov}(\varepsilon_{in}, W_{mn,i} Y_n) &= \sigma_{i,mn}^2 (I - \lambda_m W'_{mn})^{-1} X_n \beta_m \end{aligned} \quad (11)$$

where  $\sigma_{i,mn}^2$  is the variance of the error term of individual  $i$  in model  $m$ . The drawback of this predictor is that  $\sigma_{i,mn}^2$ , computed from  $\hat{\varepsilon}_{i,mn}^2$ , cannot be consistently estimated. Even though  $\Lambda_1$  is smaller than  $\Lambda_2$ , we show in section 5 below, using Monte Carlo experiments, that the first predictor should be preferred over the second to construct the  $J$  test. In the following, we will thus only consider the first predictor for the sake of clarity.

Let us denote  $\hat{Y}_n^M = [\hat{y}_n^2, \dots, \hat{y}_n^M]$ , the  $n \times (M - 1)$  matrix containing the  $M - 1$  predicted alternative models using the first predictor. To construct the  $J$  test, we augment the model under the null (model 6) with the predictors  $\hat{Y}_n^M$ :

$$Y_n = X_{1n} \beta_1 + \lambda_1 W_{1n} Y_n + \hat{Y}_n^M \delta + \varepsilon_{1n}, \quad (12)$$

where  $\delta$  is a  $(M - 1) \times 1$  vector of parameters:  $\delta' = [\delta_2, \dots, \delta_M]$ . If the model under the null is the correct one,  $\delta$  should not significantly differ from 0. This is the  $J$  test.

---

<sup>9</sup>Finally, Kelejian and Prucha (2007) propose the full information predictor but it is not more efficient than the second predictor for the SAR specification.

Our  $J$  test procedure based on the RGMM estimation procedure is summarized in the following steps:

**Step 1:** Estimate the parameters  $\beta_m$  and  $\lambda_m$  for  $m = 2, \dots, M$  by the RGMM described in section 2 and Appendix A.

**Step 2:** Estimate  $\hat{y}_n^m$  for  $m = 2, \dots, M$  and form  $\hat{Y}_n^M = [\hat{y}_n^2, \dots, \hat{y}_n^M]$ .

**Step 3:** Construct the augmented model  $Y_n = X_n\beta_1 + \lambda_1 W_{1n}Y_n + \hat{Y}_n^M\delta + \varepsilon_{1n}$  as proposed by Davidson and MacKinnon (1981, p.782). Let us note that the considered interaction matrices should represent different interaction scheme, i.e. should not be too correlated, to avoid multicollinearity between the predicted values  $\hat{Y}_n^M$ . LeSage and Pace (2014) propose to rely on correlation matrices between interaction matrices based on the spatial lags  $W_1u, W_2u, \dots, W_Mu$ , where  $u$  is a  $n \times 1$  vector of values drawn from a standard normal distribution. Ideally, we would like to observe low correlation between these spatial lags.

**Step 4:** Estimate the augmented model by RGMM using the same set of instruments as the model under the null and construct the  $J$  test as follows:

Define  $\gamma_1 = (\beta_1', \lambda_1, \delta')'$ . Let  $R = (0_{(M-1) \times (k_1+1)}, I_{(M-1)})$  so that  $R\gamma_1 = 0$ . Then, the  $J$  statistic defined as a Wald statistic is:

$$J_{1,n} = (R\hat{\gamma}_1)' \left[ R(\hat{D}_n'(\hat{\Omega}_n^{-1})\hat{D}_n)^{-1} R' \right]^{-1} (R\hat{\gamma}_1) \sim \chi_{(M-1)}^2. \quad (13)$$

where  $\hat{D}_n'$  and  $\hat{\Omega}_n^{-1}$  are respectively the gradient and the inverse of the covariance matrix of the augmented RGMM estimation procedure. In this paper, we follow Davidson and MacKinnon (1985) who suggest to use the variance-covariance matrix of the augmented specification.<sup>10</sup>

### 3.3. Statistical inference based on bootstrap

Since the work of Godfrey and Pesaran (1983); Godfrey (1998) and Davidson and MacKinnon (2002), we know that bootstrapped-based inference for the  $J$  test works much better. The reason is that the  $J$  test statistic does not have a zero mean in finite samples. It generally has a positive mean, which can

---

<sup>10</sup>Alternatively, Davidson and MacKinnon (1985) also propose to use the estimated covariance matrix coming from the estimation of the model under the null hypothesis. However, the Monte Carlo simulations we performed indicate poor performance of this alternative method.

be quite large when the sample size is small, the model being tested fits poorly, and/or there are several variables in the alternative that are not in model under the null (see Davidson and MacKinnon, 2002).

In the spatial econometrics literature, Burridge and Fingleton (2010) implement a semi-parametric bootstrap approach to the  $J$  test applied to the choice of the interaction matrix.<sup>11</sup> Besides, Jin and Lee (2015) and Yang (2015) formally prove the consistency of the bootstrap approach in spatial econometrics for linear-quadratic forms of residuals test-statistics, such as Moran's I, Lagrange multiplier statistics or  $J$  tests. In this paper, to account for heteroskedasticity, we follow Hagemann (2012) and implement the wild bootstrap proposed Liu (1988) and further developed by Mammen (1993).<sup>12</sup> Suppose  $\mathcal{F} := \{X_n, W_{mn}, m = 1, \dots, M \cup m^*; \}$ , where  $m^*$  refers to the interaction matrix associated with the correct model  $m^*$ .<sup>13</sup> The used wild bootstrap procedure consists in perturbing the residuals with *iid* copies  $\eta_1, \eta_2, \dots, \eta_n$  of a random variable  $\eta$  with  $\mathbf{E}(\eta|y, \mathcal{F}) = 0$ ,  $\mathbf{E}(\eta^2|y, \mathcal{F}) = 1$  and  $\mathbf{E}(|\eta|^{2+\zeta}|y, \mathcal{F}) < \infty$  for some  $\zeta > 0$ . Hagemann (2012), following Davidson and Flachaire (2008), recommends the use of a Rademacher distribution for  $\eta$  that takes on the value 1 with probability 0.5 and  $-1$  with probability 0.5. Finally, let  $H$  be a diagonal matrix having as diagonal elements the terms  $\eta_1, \eta_2, \dots, \eta_n$ .

Considering model as the model under the null, the bootstrap-based inference for the  $J$  test is constructed as follows:

- i) Compute the  $J$  test associated to (6) using the methodology presented above and label it  $\widehat{J}_{1,n}$ .
- ii) Estimate model (6) to get  $\widehat{\beta}_1$ ,  $\widehat{\lambda}_1$  and the residuals  $\widehat{\varepsilon}_{1n}$
- iii) Generate  $\varepsilon_{1n}^* = H\widehat{\varepsilon}_{1n}$
- iv) Generate the bootstrapped sample  $Y_n^* = (I_n - \widehat{\lambda}_1 W_{1n})^{-1}[X_n \widehat{\beta}_1 + \varepsilon_{1n}^*]$
- v) Compute the bootstrapped  $J$  test for the model (6) using the bootstrapped dependent variable  $Y_n^*$

---

<sup>11</sup>Burridge and Fingleton (2010) consider a homoskedastic error term, a first step estimation procedure based on maximum likelihood and an augmented model estimated by IV. We differ from that paper by allowing unknown heteroskedasticity in the error term and estimating all models by Robust GMM.

<sup>12</sup>According to MacKinnon (2002), the appropriate bootstrap methods to deal with heteroskedasticity of unknown form are the wild bootstrap and the pairs bootstrap proposed by Freedman (1981). MacKinnon (2002) and Flachaire (2005) compare the performance of these two bootstrap techniques and show that the wild bootstrap, using the Rademacher distribution developed in Davidson and Flachaire (2008), outperforms the pairs bootstrap in terms of error rejection probability and power.

<sup>13</sup>According to Hagemann (2012), this design matrix may be observed or unobserved.

and label it  $J_{1,n}^*$ .<sup>14</sup>

Repeat operations *iii)* to *v)*  $B$  times, each time with a new realization of  $H$ . Reject the null hypothesis at the  $\alpha$  threshold if  $\widehat{J}_{1,n}$  is larger than  $c_{1-\alpha}^*$ , the  $1 - \alpha$  empirical quantile of the distribution of  $J_{1,n}^*$ .

#### 4. The $MJ$ -test

The  $J$  test presented above is a procedure for testing a null model against non-nested alternatives. It is known that tests of non-nested models may suffer from a problem of model selection. To overcome this drawback, Hagemann (2012) suggests a simple test, the  $MJ$  (Minimum  $J$ ) statistic which allows to always select a model between competing alternatives. The  $MJ$  statistic consists in selecting the model with the smallest  $J$  statistic value. In this paper, we extend the  $MJ$  test procedure to the selection of the interaction matrix in spatial autoregressive models.

Assume one wishes to find the most appropriate interaction matrix among a set of  $M$  spatial autoregressive models.

$$Y_n = X_n \beta_m + \lambda_m W_{mn} Y_n + \varepsilon_{mn}, \quad m \in \mathcal{M} = \{1, \dots, M\}, \quad (14)$$

with  $\varepsilon_{mn}$ , the  $n \times 1$  vector of independent error terms with  $\varepsilon_{i,mn}$  characterized by a zero mean and a variance  $\sigma_{i,mn}^2$ ,  $i = 1, \dots, n$  and  $\theta_m = (\lambda_m, \beta_m')'$  is the vector of regression parameters to be estimated for  $m = 1, \dots, M$ . The  $MJ$  test procedure is the following:

**Step 1:** For each  $m \in \mathcal{M}$ , compute associated  $J$  test, corresponding to testing model  $m$  against the  $M - 1$  alternatives non-nested hypotheses.

For this, estimate the augmented model by RGMM:

$$Y_n = X_{mn} \beta_m + \lambda_m W_{mn} Y_n + \sum_{l \in \mathcal{M} \setminus \{m\}} \delta_l \widehat{y}_n^l + \varepsilon_{mn}. \quad (15)$$

Construct then  $\widehat{J}_{m,n}$ , the  $J$ -test associated to model  $m$ , as a test of joint significance of  $\delta_l, l \in \mathcal{M} \setminus \{m\}$ .

---

<sup>14</sup>The bootstrapped dependent variable is also used to estimate the models under the alternative and obtain the associated predicted values.

**Step 2:** Let  $\mathcal{J}_n = \{\hat{J}_{m,n} : m \in \mathcal{M}\}$  and define the least  $J$  statistic:  $MJ_n = \min \mathcal{J}_n$ . This step solves the decision problem inherent to the  $J$  test since the preferred model in the set of the  $M$  considered specifications will be the one with the smallest value of  $\hat{J}_{m,n}$ ,  $m = 1, \dots, M$ .<sup>15</sup>

## 5. Monte Carlo experiments

The objective of these Monte Carlo experiments is twofold. Firstly, we aim at comparing the performance of the RGMME with respect to TSLS in terms of efficiency. since the former accounts for more information than the latter. Secondly, we wish to assess the small sample properties of the  $J$  and  $MJ$  tests relying on the first or the second predictor developed in section 3.2, within the RGMM framework. Even though the information content of the second predictor (based on the structural form) is larger than that of the first predictor (based on the reduced form), we have seen that the former requires the estimation of heteroskedastic variances for which no consistent estimator exists.

### 5.1. Comparison of RGMME and TSLSE

To compare the efficiency of the RGMME with respect to the TSLSE, we consider the following DGP:

$$Y_n = \lambda W_n Y_n + \iota_n \beta_0 + X_{1n} \beta_1 + X_{2n} \beta_2 + \varepsilon_n. \quad (16)$$

where  $\iota_n$  is the unit vector,  $X_{1n} \sim U(0, 10)$ ,  $X_{2n} \sim N(0, 1)$  and the associated vector of parameters  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  are all set equal to 1. In these simulations, the regressors are kept constant across replications. The error term is assumed heteroskedastic and non-normally distributed. More precisely, it is designed in the following way:

$$\varepsilon_{in} = \bar{\sigma} \left( \frac{\nu_{in} - 2}{\sqrt{2}} \right) X_{i,1n}, \quad (17)$$

where  $\nu_{in}$  follows a  $\Gamma(2, 1)$  distribution. Besides, to make simulation results comparable across different values of parameters, we set the average dispersion of the error term ( $\bar{\sigma}$ ) so that the signal to noise ratio is maintained to a given value. This signal to noise ratio is computed as the ratio of the explained variance over the total variance, ignoring spatial autocorrelation (see Lee, 2007; Liu et al., 2010). For this comparison

---

<sup>15</sup>Hagemann (2012) also proposes to test whether the correct model is part of the considered ones. However, in this paper, we focus only on selecting the best interaction matrix among a set of possible choices.

of estimators, we consider signal to noise values of 0.3 and 0.7. We consider a five nearest neighbors interaction matrix constructed from a random set of coordinates. Further, the matrix is normalized by its spectral radius. The parameter  $\lambda$  takes on the value of  $[-0.7, -0.5, -0.3, 0.3, 0.5, 0.7]$  to capture low, moderate and high positive and negative cross-sectional dependence. Two sample sizes,  $n = 60$  and  $n = 100$ , are considered. Finally, for each case, we perform 1000 replications.

Table 1 summarizes the comparison between RGMM and TSLS estimators for the two considered sample sizes. The TSLS is constructed using the best instruments developed by Lee (2003), collected in the matrix  $H = [W(I_n - \hat{\lambda}W_n)^{-1}X_n\hat{\beta}, X_n]_{LI}$ , where  $LI$  stands for linearly independent columns.<sup>16</sup> The matrix of instruments used for the RGMM procedure is presented in Appendix A. To compare the two estimators, we rely on the definition of bias and efficiency of Kelejian and Prucha (1998), based on the empirical distribution of the estimator obtained from the 1000 replications. Specifically, the bias is defined as the absolute difference between the median of the empirical distribution and the true parameter value, while the efficiency, measured as a RMSE, is shown in (18), with  $IQ$  the interquartile range.

$$RMSE = \left[ Bias^2 + (IQ/1.35)^2 \right]^{1/2} \quad (18)$$

From the bias perspective, we do not observe significant difference between the two estimators. Table 1 shows that the bias is low for both of them, even for  $n = 60$ . However, we note that when the signal to noise ratio is low (0.3), the RGMM of  $\lambda$  is more efficient than its TSLS counterpart. This is especially true for small sample size where the RMSE of TSLS for  $\lambda$  is between 28% and 35% higher. This difference diminishes to 2-7% when the signal to noise ratio increases to 0.7. When the sample size increases, performance of the TSLS for  $\lambda$  get closer to those of RGMM, and as expected, results are even closer when the signal to noise ratio is higher. Finally, we note that for the other parameters, bias and efficiency are similar across estimation methods.

---

<sup>16</sup>As this set of instruments involves the parameters  $\lambda$  and  $\beta$ , we adopt the following procedure: we first estimate  $\lambda_0$  and  $\beta_0$  by using the instrument matrix  $H = [X_n, W_n X_n, W_n^2 X_n]_{LI}$ . Once we have an estimated values for the parameters, we can consider the matrix of the best instruments :  $H(\hat{\theta}) = [W(I_n - \hat{\lambda}W_n)^{-1}X_n\hat{\beta}, X_n]_{LI}$ . We then iterate between TSLS estimation and construction of the matrix of best instruments until we reach convergence. The selected convergence criterion is  $\sum_{j=1}^{K+1} |\hat{\theta}_{j,s+1} - \hat{\theta}_{j,s}| < 0.0001$ , where  $\hat{\theta}_{j,s}$  is the estimator of  $\theta_{0,j}$  obtained after  $s$  iterations.



Table 1: Comparison of RGMM and TSLS estimators,  $W_n = 5$  nn

$n = 60$													
Signal to noise = 0.3													
		$\lambda = -0.7$		$\lambda = -0.5$		$\lambda = -0.3$		$\lambda = 0.3$		$\lambda = 0.5$		$\lambda = 0.7$	
		RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS
$\lambda$	Bias	0.006	0.016	0.003	0.014	0.002	0.013	0.005	0.013	0.010	0.008	0.010	0.011
	RMSE	0.267	0.342	0.268	0.343	0.261	0.339	0.207	0.267	0.176	0.229	0.140	0.188
$\beta_0$	Bias	0.046	0.054	0.042	0.054	0.013	0.050	0.084	0.154	0.121	0.162	0.210	0.258
	RMSE	1.353	1.667	1.441	1.773	1.564	1.899	1.942	2.473	2.286	2.848	2.918	3.790
$\beta_1$	Bias	0.028	0.025	0.027	0.026	0.026	0.028	0.022	0.024	0.021	0.022	0.019	0.016
	RMSE	0.203	0.202	0.200	0.201	0.200	0.203	0.200	0.202	0.199	0.201	0.199	0.200
$\beta_2$	Bias	0.068	0.066	0.071	0.068	0.072	0.071	0.071	0.068	0.070	0.064	0.067	0.058
	RMSE	0.430	0.439	0.430	0.438	0.428	0.437	0.426	0.431	0.425	0.427	0.424	0.436
Signal to noise = 0.7													
		RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS
$\lambda$	Bias	0.001	0.004	0.002	0.005	0.002	0.006	0.001	0.004	0.003	0.002	0.003	0.002
	RMSE	0.142	0.145	0.141	0.144	0.138	0.144	0.110	0.114	0.092	0.096	0.074	0.079
$\beta_0$	Bias	0.003	0.013	0.019	0.021	0.033	0.019	0.040	0.050	0.067	0.052	0.082	0.108
	RMSE	0.694	0.722	0.734	0.768	0.774	0.821	0.988	1.052	1.171	1.250	1.512	1.585
$\beta_1$	Bias	0.010	0.009	0.010	0.008	0.009	0.008	0.008	0.008	0.008	0.007	0.007	0.006
	RMSE	0.088	0.087	0.087	0.087	0.087	0.087	0.085	0.086	0.086	0.085	0.084	0.085
$\beta_2$	Bias	0.027	0.027	0.027	0.027	0.028	0.026	0.027	0.026	0.025	0.025	0.024	0.022
	RMSE	0.188	0.189	0.187	0.189	0.188	0.187	0.184	0.182	0.181	0.181	0.181	0.183
$n = 100$													
Signal to noise = 0.3													
		$\lambda = -0.7$		$\lambda = -0.5$		$\lambda = -0.3$		$\lambda = 0.3$		$\lambda = 0.5$		$\lambda = 0.7$	
		RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS
$\lambda$	Bias	0.004	0.018	0.006	0.016	0.005	0.014	0.001	0.006	0.004	0.005	0.002	0.004
	RMSE	0.188	0.227	0.189	0.232	0.187	0.224	0.142	0.186	0.122	0.163	0.094	0.131
$\beta_0$	Bias	0.028	0.119	0.006	0.092	0.002	0.067	0.056	0.069	0.064	0.103	0.133	0.007
	RMSE	0.893	1.044	0.960	1.176	1.056	1.265	1.289	1.649	1.514	1.964	1.887	2.591
$\beta_1$	Bias	0.010	0.005	0.011	0.007	0.012	0.008	0.013	0.010	0.012	0.010	0.011	0.008
	RMSE	0.102	0.106	0.102	0.105	0.102	0.104	0.101	0.102	0.103	0.102	0.103	0.100
$\beta_2$	Bias	0.004	0.001	0.003	0.001	0.003	0.001	0.004	0.002	0.004	0.002	0.004	0.000
	RMSE	0.255	0.263	0.253	0.260	0.251	0.258	0.250	0.254	0.253	0.256	0.252	0.255
Signal to noise = 0.7													
		RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS	RGMM	TSLS
$\lambda$	Bias	0.007	0.008	0.006	0.007	0.004	0.006	0.001	0.004	0.000	0.004	0.000	0.002
	RMSE	0.094	0.097	0.097	0.099	0.095	0.097	0.075	0.080	0.066	0.070	0.053	0.056
$\beta_0$	Bias	0.027	0.040	0.017	0.037	0.009	0.032	0.010	0.030	0.004	0.035	0.034	0.006
	RMSE	0.449	0.467	0.488	0.511	0.529	0.551	0.687	0.717	0.797	0.850	1.040	1.145
$\beta_1$	Bias	0.003	0.001	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002
	RMSE	0.045	0.045	0.045	0.045	0.044	0.044	0.043	0.042	0.043	0.042	0.043	0.043
$\beta_2$	Bias	0.001	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.001	0.000
	RMSE	0.112	0.113	0.110	0.110	0.110	0.109	0.108	0.109	0.108	0.109	0.108	0.110

## 5.2. Performance of the 2 predictors and $MJ$ test

In this section, we assess the performance of the 2 predictors presented in Section 3 and the  $MJ$  test. For this, we consider 3 competing SAR models, shown in (19),

$$\begin{aligned} M_1 : Y_n &= \lambda_1 W_{1n} Y_n + X_n \beta + \varepsilon_{1n} \\ M_2 : Y_n &= \lambda_2 W_{2n} Y_n + X_n \beta + \varepsilon_{2n} \\ M_3 : Y_n &= \lambda_3 W_{3n} Y_n + X_n \beta + \varepsilon_{3n}, \end{aligned} \tag{19}$$

where the matrix of explanatory variables  $X_n$ , the values for  $\beta$  and  $\lambda$ , the number of replications and the error term  $\varepsilon_{1n}$ ,  $\varepsilon_{2n}$  and  $\varepsilon_{3n}$  are defined identically to section 5.1. In these Monte Carlo, we use a signal to noise value of 0.7. The matrices  $W_{1n}$ ,  $W_{2n}$  and  $W_{3n}$  are constructed using three different sets of normally drawn random coordinates, to guarantee they model different interaction schemes. In the experiments, we rely on interaction matrices based on 5, 10 and 20 nearest neighbors (nn).

In the simulations,  $M_1$  is assumed to be the DGP and  $W_{1n}$  is relabeled  $W_{correct}$  in Tables 2 and 3 to make this clear. As such, the rejection rate of the  $J$  test associated to  $M_1$  measures its size, while the rejection rates for the  $J$  statistic associated to  $M_2$  and  $M_3$  represent a measure of its power. As we consider 3 models, the  $J$  test is constructed using a Wald statistic (see Step 3 of section 3).

Table 2: Comparison of the  $J$  and  $MJ$  tests performance,  $n = 60$

$\lambda$	First Predictor						Second Predictor							
	% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)			% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)		
		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$
	$W_{correct}: 5 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 5 \text{ nn}$													
-0.7	98.7	4.6	92.4	90.6	8.7	96.4	95.7	84.7	2.4	48.9	49.8	11.1	73	72.6
-0.5	94.4	5.1	71.6	70.5	8.6	81.1	80	79.2	3	38.5	37.8	12.4	64.9	65.1
-0.3	73	5.3	33.9	32.6	8.9	43.4	43.4	67.4	3.9	25.3	25	14	47.1	48.3
0.3	82.1	4.8	55.4	55.5	8.8	65.6	62.8	49.5	4.1	11.5	10.8	14.6	26.5	28.8
0.5	97	5	91.7	89.9	8.8	95	94	52.8	3.8	14.8	15.5	12.5	31.6	35.5
0.7	99.5	4.9	99.1	98.5	9	99.4	99.2	60.3	2.6	15.7	17.3	10.4	37.5	47.9
	$W_{correct}: 5 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 5 \text{ nn}$													
-0.7	98.4	5.3	91.4	91.7	8.9	95.4	95.7	83.3	2.6	49	43.8	13.6	72	71.5
-0.5	92.4	5.5	68.1	72.4	8.9	79.2	82.5	77.7	3.2	38.6	35.5	13.8	61.9	63.1
-0.3	73.8	5.1	30.8	33.6	8.9	42.4	43.6	65.3	4	25	23.3	15.2	46.2	48.4
0.3	83.3	4.9	55.9	58.2	8.8	66.7	65.4	48.1	3.9	11.8	11.1	16.2	28.3	31.5
0.5	97.6	4.9	92.1	92.2	8.9	95.4	95.4	52.8	2.9	14.4	13.9	13.7	32.8	34.1
0.7	99.5	4.9	99.1	99	8.9	99.7	99.8	61.4	2.4	16	14.8	11.6	39.7	44
	$W_{correct}: 5 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 5 \text{ nn}$													
-0.7	98.3	5.5	91.4	90.5	8.7	96	94.7	80.6	2.8	49.6	41.4	19.1	71.8	78.5
-0.5	92.9	5.5	69.8	69	9.3	80.1	79	76.4	3.5	38.4	33.4	20.3	63.1	69.8
-0.3	73.8	5.3	31.6	31.5	9.3	41.9	39.6	63.1	3.8	25.9	20.4	23.4	46.2	54.1
0.3	82.2	4.8	55.8	55.4	8.1	66.4	62.4	43.9	4.6	11.4	12.1	23.2	27.7	38.6
0.5	96.1	5.4	92	89.3	8.3	95.4	94.4	50.1	3.2	14.7	15	20.1	32.6	44.2

Continued on next page

Table 2 – Continued from previous page

$\lambda$	First Predictor							Second Predictor						
	% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)			% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)		
		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$
0.7	99.5	5.4	99	98.9	8.8	99.4	99.2	58.4	2	16.4	15.8	16.1	39.3	52.8
	$W_{correct}: 5 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 10 \text{ nn}$													
-0.7	98.4	4.8	90	89.9	9.1	96.3	95.1	86.1	3.1	48.8	49.3	12	72.7	73
-0.5	93.3	4.9	70.2	70.6	9.2	80.2	79.3	80.3	3.4	38.5	39.4	13	63.4	64.8
-0.3	71.9	4.7	31.9	34	9.3	44.7	44.3	69.2	4.2	25.7	26.1	14.3	47.6	48.9
0.3	82.1	4.9	58.3	53.9	9.4	68.4	62.3	51.8	5.1	14.8	12.7	13.7	32.4	31.1
0.5	97.6	4.9	92	89.2	9.4	95.9	93.8	58.4	3.5	17.8	16.2	12.5	39	37.2
0.7	99.7	5.1	99.2	98.2	9.6	99.8	99.1	71.2	2.9	21.1	20.1	9.8	53.2	50.6
	$W_{correct}: 5 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 10 \text{ nn}$													
-0.7	98.6	4.7	88.5	91.5	9.3	94.8	95.1	84.1	3.2	48.1	45.9	14.7	70.5	72.1
-0.5	91.8	4.7	68.3	73	9.6	78.6	82.8	77.5	3.5	39.1	36.8	15.6	62.6	62.5
-0.3	73.7	4.7	30.3	34.6	9.4	43.6	44.4	66.4	4.3	24.5	24.1	16.6	46.5	48.2
0.3	82.5	4.7	57.3	56.2	9.4	68.4	65.7	50.2	4.5	14.2	12.5	16.4	32.3	32.6
0.5	97.5	5	92.3	91.4	8.9	95.8	95.1	55.8	3.7	17.8	13.7	13.3	39.6	34.2
0.7	99.6	4.9	99.1	98.7	8.8	99.8	99.4	69	2.8	20.9	16	11.3	54.8	47.3
	$W_{correct}: 5 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 10 \text{ nn};$													
-0.7	98.6	5	89.8	90.4	8.8	95.3	94.6	81.2	2.9	49.1	42.1	19.7	71.9	78.9
-0.5	92.9	5.4	70.1	69.2	9.2	80.7	79.5	74.4	3.4	38.8	34.3	21	63.2	70.1
-0.3	72.2	5.4	30.1	31.8	9.3	42.8	41.5	62.3	3.7	26.1	21.5	22.2	46.6	56.5
0.3	81.6	5.4	58.4	54.2	9.1	68.6	61.3	46.2	4.6	13.9	13.3	24	33.2	40.6
0.5	97.1	5.7	92.2	89.4	8.5	95.9	94	52.7	3	17.9	15.4	21	39.5	44.7
0.7	99.9	5.5	99.3	98.7	8.7	99.8	99.3	66.3	2.5	19.9	16.7	14.7	49.2	53.3
	$W_{correct}: 5 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 20 \text{ nn}$													
-0.7	98.5	4.7	89.1	89.9	8.7	94.1	95	88.9	3.7	53.2	55.6	16.4	77	76
-0.5	91.9	4.7	68	69.5	8.6	77.7	79.2	83	4.4	43.7	46.5	17.5	68.5	67.2
-0.3	72.8	4.9	30.8	33	8.8	41.5	43.1	68.3	5.1	25.4	29.5	19.3	53.8	52.5
0.3	83	4	56.7	55.5	9.4	67.4	64.1	55.2	6.3	17.4	21.7	18.7	48.3	40.1
0.5	97	4	92.3	90.1	9.3	94.8	93.9	63.2	4.8	31.8	33.8	16.7	56.6	51.2
0.7	99.7	4.3	98.7	98.6	9.5	99.4	99.2	73.8	3.8	37.4	39.6	13.6	61.6	61
	$W_{correct}: 5 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 20 \text{ nn}$													
-0.7	97.7	5	87.9	92	9.3	93.1	95.3	87.5	3.9	52.3	53.4	16.9	76.3	74.4
-0.5	91.4	4.9	65.8	73.3	9.3	76.7	82.1	82.3	4.4	43.5	45.7	18.1	66.8	67.8
-0.3	72.5	5	29.7	36.4	9.3	41.1	44.1	67.1	4.9	26.1	27.4	19.6	53.9	51.7
0.3	82.4	4.9	56.3	57.2	9.2	67.8	65.8	53	6.1	19.1	19.6	21.4	48	41.9
0.5	97	4.6	92.7	91.4	9	94.9	94.8	64.6	5.5	32.6	32.7	18	57.1	50.8
0.7	99.7	4.7	98.7	98.9	8.6	99.3	99.6	73.8	3.9	38.2	37.1	16.5	62.1	59.3
	$W_{correct}: 5 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 20 \text{ nn}$													
-0.7	98.2	4.2	89.1	89.7	8.3	93.7	94.9	85.9	3.3	53	50.1	22.6	76.2	79.9
-0.5	92.5	4.3	66.4	68.3	8.8	77.8	78.3	79.2	4	43.9	40.8	24.8	67.7	72.5
-0.3	69.8	4.4	29.6	33.7	8.8	41.1	41.8	63.4	4.9	26.1	24.3	27.5	53.6	59
0.3	81.9	4.4	57.3	55.3	8.3	68.8	63.3	49.6	6	18.9	19.7	27.7	47.8	49.2
0.5	96.4	4.7	92.2	90.2	8.7	94.9	93.5	60.8	4.7	32.2	33.5	24.4	57.4	58.7
0.7	100	4.9	98.7	98.8	8.5	99.4	99.3	72.4	2.8	38	37.5	18.1	62.3	64.5
	$W_{correct}: 10 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 5 \text{ nn}$													
-0.7	80.3	4	45.8	44.8	8.4	59.2	56.5	81	3.3	34.8	32.4	12.7	61.7	61.6
-0.5	67.4	4.2	29.5	26.1	8.4	39.1	38.4	69.3	3.5	23.7	21.6	13.5	48.6	48.7
-0.3	50	4.3	12	10.3	8.2	20.3	19.7	55.5	4	14.3	12.5	14.1	33.6	32.7
0.3	57.7	4.4	22.8	23.7	8.1	30.2	31.2	40	4.4	6.9	7.6	15.3	22	20
0.5	81.2	4.7	53.6	55.6	8.4	64.7	63.6	48.9	4.2	14.6	14.2	14.4	31	29.9
0.7	93.4	4.8	79.9	80.1	8.2	84.9	84.4	61.9	4	24.5	21.1	12.6	46.3	43.5
	$W_{correct}: 10 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 5 \text{ nn}$													
-0.7	79.1	4.5	46	43.8	8.1	58.2	54.9	79.6	3.5	36.1	31	15.7	62.7	63
-0.5	67.6	4.8	29.3	24.6	8.4	39	35.9	68.1	3.4	24.4	20.5	16.2	49.2	48.1
-0.3	49.8	4.7	12.9	11	8.4	20.9	18.3	53.9	3.8	15	11.6	16.2	33.9	33.8
0.3	60.9	5.1	23.5	25.8	8.4	30.5	33	37	4.2	7	7.2	16.9	23	22
0.5	83	5.2	54.5	58.1	8.5	64.6	66.6	48.7	3.6	14	13.2	15.8	30.1	31.2
0.7	93.3	5.1	80.2	83.1	8.5	84.8	87.8	61	2.6	25.7	21.5	14.5	45.8	45
	$W_{correct}: 10 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 5 \text{ nn}$													
-0.7	80.3	5.7	45.4	43.4	8.2	57.7	55.2	76.6	3.8	34.9	29.4	21.5	61.2	68.6
-0.5	66	5.5	27.8	26.4	8.1	40.1	36.9	65.5	3.8	23.7	18.6	23	48.3	57.4
-0.3	49.5	5.8	11.9	10.8	8.5	19.5	18.7	50.8	4.5	14.1	10.7	23.7	33.5	41.1
0.3	59.5	5.4	22.8	24.4	8.9	31	29.8	34.4	4.5	7.5	7.1	24.7	22.4	30.1
0.5	81.5	5.5	54.2	55.8	8.9	65.3	64.2	46.4	4.2	14.6	12.2	23.3	31.4	41.3

Continued on next page

Table 2 – Continued from previous page

$\lambda$	First Predictor							Second Predictor						
	% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)			% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)		
		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$
0.7	93.2	5.5	80.3	79.9	9	85.2	84.6	60.4	2.9	24.4	21.7	20.5	46.3	56.2
	$W_{correct}: 10 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 10 \text{ nn}$													
-0.7	80.6	4.4	44.3	46.8	9.4	58.8	56.4	80.5	4	29.7	30.9	14.1	59.9	61.1
-0.5	66.9	4.7	26.2	27.8	9.3	40.1	38.9	69	4.2	21.3	20.9	14.5	47.3	47.9
-0.3	50.4	4.7	11.9	11.6	9.3	22.4	20.7	54.3	4.6	12.1	11.6	14.6	34.1	33.1
0.3	58.5	4.8	22.8	22.8	9.1	34.3	31	41	5.1	8.4	7.5	14.8	23.4	21.6
0.5	80.1	5.2	53.8	55	8.8	67.2	63.3	51.2	4.9	14.7	13.2	14.1	33.9	29.9
0.7	91.8	5.3	81.3	78.8	9.2	87.8	83.5	63.5	4.6	23.7	21.6	13.6	48.9	43.1
	$W_{correct}: 10 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 10 \text{ nn}$													
-0.7	79.7	5.5	45.1	45.1	9.3	58.2	55.6	78	3.9	30.3	29.6	15.8	59.8	62.4
-0.5	67.4	5.5	27.8	27.1	9.4	41	37.6	68.1	4.2	20.8	19.5	16.4	47.4	48.2
-0.3	50.7	5.6	12.3	10.8	9.5	23	18.5	52.8	4	11.6	10.7	16.7	32.8	34.5
0.3	60.2	5.6	23.8	25.1	9.4	34	33.1	39.9	4.4	8.3	6.9	17.3	23.4	23.8
0.5	82.4	5.5	55.1	56.5	9.1	67.5	65.4	49.6	4.1	14.5	12.5	15.9	34.5	30.8
0.7	92.3	5.3	81.7	80.9	9.1	88.1	87.2	63.2	3.6	24.3	21.7	14	47.4	45.5
	$W_{correct}: 10 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 10 \text{ nn}$													
-0.7	79.2	5.1	44.7	44.9	8.9	57.4	57.1	75.6	3.9	30.5	27.4	23.1	59	67.8
-0.5	67.1	5.8	26.1	27.7	9.1	39.6	36.3	65.2	4.3	20.8	17.8	24.1	46.9	55.4
-0.3	48.6	5.7	10.9	11.6	9.4	21.3	18.9	50.3	4.4	11.5	10.7	23.9	31.5	40.9
0.3	60.3	5.2	24.8	24.7	9.8	35.4	31.3	36.6	4.1	9.1	6.5	24.4	24	32
0.5	80	5	55.9	54.5	10	68.3	63.9	47	3.8	14.1	11.1	23.4	35.1	41
0.7	91.7	5.2	81.8	78.8	9.6	88.2	83.7	61.9	3.2	23.9	21.3	21.7	48.8	56.7
	$W_{correct}: 10 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 20 \text{ nn}$													
-0.7	81.4	5.2	44.6	46.4	9.1	59.2	56.8	79.9	3.6	29.6	35.3	17.4	64.3	65.9
-0.5	67.1	5.1	25.4	27.5	9.3	41.3	38.7	69.3	4.2	20.7	23.5	18.2	52.2	50.1
-0.3	51	5.4	10.1	11.7	9	21.7	20	53.4	5.1	11.6	12.3	19.3	38.3	33.5
0.3	59.7	5	23.6	24.2	9.1	35.2	33.1	42.5	6.6	7.1	8.9	20.9	31.2	26.1
0.5	81.6	4.5	57.4	57.8	9.3	67.9	65.3	53	6.5	16.3	19.3	19.6	42.7	37.2
0.7	93.6	4.5	82	81.2	9.2	88	85	61.8	5.1	24.4	26.8	18	51.3	47.5
	$W_{correct}: 10 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 20 \text{ nn}$													
-0.7	79.2	4.9	44.3	45.1	9.3	58.6	55.9	78.8	3.9	29.8	33.9	18.6	62.9	66
-0.5	66.4	4.7	25.8	27.3	9.2	40.5	37.3	65.9	3.9	20.5	22.4	19.9	50.5	50.2
-0.3	50.7	5.1	10.3	11.2	9.4	21.7	19.1	51.1	4.3	12.1	11.6	20.2	37.7	33.2
0.3	59.9	4.9	23.9	25	9.4	36	34.2	40.8	6.2	7.6	8.6	22.6	32	26.6
0.5	82.3	4.8	58.5	59	9.4	69.2	67.8	52	6.4	16.5	18.8	20.6	44.6	39.4
0.7	92.9	4.8	81.7	82.8	9.1	88.5	87.9	61.5	4.6	25.2	27.4	18.1	51.6	49.5
	$W_{correct}: 10 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 20 \text{ nn}$													
-0.7	80.2	4.3	42.6	44.2	9.4	58.7	57.8	75.2	3.9	30.2	31	24.9	63.8	71.6
-0.5	67.3	4	24.9	26.7	9.4	40.9	37.7	63.2	4.3	20.6	18.9	26.5	52.3	57.8
-0.3	49.9	4.4	10.1	11.8	9.9	22.3	19.2	45.9	4.3	11.1	11	27.5	37.6	42.5
0.3	60.6	5	24.9	24.8	9.8	37	32.2	37.1	5.5	7.9	7.5	29.7	31.3	35.1
0.5	81.9	4.8	58.5	58.2	9	69.1	67.3	50.5	5	17.4	17.1	28.5	44.6	47.9
0.7	92.1	5	81.8	81.3	8.6	88.4	85.7	59.4	4.1	25.5	26.7	26.3	53.2	59.9

Table 3: Comparison of the  $J$  and  $MJ$  tests performance,  $n = 100$ 

$\lambda$	First Predictor						Second Predictor							
	% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)			% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)		
		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$
	$W_{correct}: 5 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 5 \text{ nn}$													
-0.7	99.8	4.4	99.6	99.5	6.2	99.8	99.7	83.9	3.5	42	42.5	13.1	67	68
-0.5	99.3	4.4	96.9	97.2	6.1	98.9	98.4	79.3	4	36.9	38.3	14.2	60.5	60.8
-0.3	94.5	4.6	74.5	73.6	6.1	81.7	81.3	71.4	4.4	29	28.5	14.4	50.7	50.4
0.3	98	4.6	92.3	92.9	6.7	95	94.5	47.3	3.6	12.9	14.1	14.3	28.4	28.2
0.5	99.7	4.9	99.7	99.7	6.6	99.7	99.9	54.3	3.1	16.8	18.7	11.4	33.4	35.4
0.7	99.7	5.1	99.7	99.8	7.1	99.7	99.8	69.4	2.3	26.6	28.1	9.5	49.1	52.1
	$W_{correct}: 5 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 5 \text{ nn}$													
-0.7	100	4.4	99.6	99.6	6.5	100	99.9	85.2	2.6	43.4	44	10.8	67.7	68.4
-0.5	99.2	4.6	96.8	96.4	6.6	98.9	97.8	81.3	3	36.2	38.1	11.1	60.6	62.4

Continued on next page

Table 3 – Continued from previous page

First Predictor							Second Predictor							
$\lambda$	% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)			% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)		
		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$
-0.3	93.7	4.4	74.7	73.7	6.5	80.8	80.7	72.8	3.2	29.2	26.9	13.2	50.8	51.4
0.3	98.1	4.5	92.9	93.4	6.8	94.8	95.2	48	3.4	12.7	13.2	12.5	28.1	28.2
0.5	99.8	4.4	99.9	99.7	7.1	99.9	99.8	55.2	2.6	16.9	18.5	11.2	34.2	33.4
0.7	99.7	4.8	99.7	99.8	7	99.7	99.8	70.3	2.6	26.9	27.7	8.8	49.5	49.1
$W_{correct}: 5 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 5 \text{ nn}$														
-0.7	99.8	4.4	99.6	99.5	5.8	99.9	99.9	84.9	2.5	42.4	43.3	14	67.5	69.5
-0.5	99.2	4.6	96.7	96.7	6.1	98.8	98.1	80.2	2.8	37.3	38.3	15.4	62	63.7
-0.3	95	4.3	74.2	73.1	6	80.8	78.9	69.9	3.5	29.9	27.8	16.7	50.2	53.8
0.3	98.2	4.7	92.7	93.4	5.9	95.1	95.5	47.7	3.9	13.5	12.5	14.5	28.1	29.9
0.5	99.6	4.6	99.6	99.9	6	99.6	99.9	54.1	2.6	17	16.7	11.4	34.4	33.7
0.7	99.8	4.9	99.8	99.8	6.4	99.8	99.8	69.6	1.9	27	25.3	9.7	49.2	50.9
$W_{correct}: 5 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 10 \text{ nn}$														
-0.7	100	5.5	99.8	99.7	8.1	100	100	83.8	3	40.6	42.9	13.1	65.2	67.8
-0.5	99.6	5.4	97.6	97.5	8	99.4	98.7	78.4	4.2	36.4	37.4	14.4	60.3	61.9
-0.3	94.3	5.4	73.5	75.1	8	81.7	81.5	71.3	4.5	27.8	30.4	15.9	50.6	51.2
0.3	98.2	5.5	92.2	92.3	7.6	95.1	94.6	51.5	6.1	18.4	17.2	16.1	35.9	32.5
0.5	99.9	5.4	99.9	99.9	7.6	99.9	99.9	60.5	5.1	26.9	22	12.6	49.9	39.8
0.7	100	5.7	100	100	7.5	100	100	75.9	4.4	45	34.9	9.9	69.8	57.9
$W_{correct}: 5 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 10 \text{ nn}$														
-0.7	100	5.2	99.8	99.7	6.8	100	99.9	85.6	2.3	41.6	44.5	10	66	68.8
-0.5	99.5	5	97.3	96.9	6.7	99.4	98.6	80.5	2.9	35.5	37.7	10.7	60.4	62.3
-0.3	94.1	5	73.6	74.7	6.6	81.7	81.4	72.5	4.3	27.5	29.9	13.2	51.1	50.7
0.3	98.2	5.2	92.2	92.9	7.4	94.9	95.1	52.6	5	18.8	17.4	13.5	35.4	31
0.5	99.9	4.9	99.9	100	7.2	99.9	100	62.3	4.1	26.8	22.1	11.3	50.7	39.7
0.7	100	5.2	100	100	7.4	100	100	76.8	3.8	46.4	34.4	9.1	70.3	57
$W_{correct}: 5 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 10 \text{ nn};$														
-0.7	99.9	5.5	99.8	99.7	7.8	100	99.9	84.9	2.3	41.9	42.8	12.8	65.2	69.1
-0.5	99.5	5.5	97.5	97	8	99.4	98.8	79.4	2.9	36.6	38.1	13.9	60	63.7
-0.3	95.1	5.3	73.1	74.8	8.5	81.9	80.1	70.7	3.6	28.5	29.6	16	50.6	54.3
0.3	98.1	5.5	92.7	92.8	8.3	95.1	95.2	51.9	4.3	19.3	15.5	15.5	35.4	31.8
0.5	99.9	5.3	99.9	99.9	7.9	99.9	99.9	61.8	3.2	26.7	20.7	12.9	50.4	40.6
0.7	100	5.5	100	100	7.9	100	100	77.2	2.8	45.5	33.1	11.1	69.7	58.3
$W_{correct}: 5 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 20 \text{ nn}$														
-0.7	99.9	5.7	99.8	99.7	7.2	99.9	100	84	3.9	40.4	43.4	13.3	64.8	67.8
-0.5	99.6	5.5	96.5	97.4	6.9	98.6	98.6	79	5.4	34.3	37.7	14.5	57.7	60.8
-0.3	94.5	5.3	72.6	73.8	7.1	79.7	81.4	68.8	6.2	28.7	30.7	15.3	47.3	49.7
0.3	97.4	5	92	91.4	7.1	93.8	94.3	50.2	6.1	20.3	17.5	15.7	37.9	32.7
0.5	99.7	5.3	99.8	99.8	7.4	99.8	99.8	60.9	5.1	24.9	23.6	13.1	45.5	41
0.7	99.9	5.3	100	100	7.4	100	100	74.6	3.8	34	33.5	10.7	61	57.3
$W_{correct}: 5 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 20 \text{ nn}$														
-0.7	99.9	4.7	99.7	99.7	7.3	99.9	99.9	85.6	3.4	40.8	45.7	10.5	66	68
-0.5	99.6	4.9	96.2	97.1	7.1	98.7	98.5	81	4	33.2	38	11.5	57.9	61.2
-0.3	93.5	5	73.6	74.8	7.1	80.4	81.2	70.6	5.1	27.9	29.8	14.1	47.4	50.9
0.3	98	4.9	91.9	92.6	7.1	94	94.9	52.1	6	19.5	18.2	14.5	37.4	32.1
0.5	99.7	4.9	99.8	99.8	6.9	99.8	99.8	59.9	4.4	24.1	24	13.1	47.4	41.3
0.7	99.9	5.2	100	100	6.9	100	100	74.9	3.5	34	33.2	11	61.2	56.9
$W_{correct}: 5 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 20 \text{ nn}$														
-0.7	99.9	4.8	99.9	99.6	6.2	100	99.8	84.6	3.5	41.2	44.8	14	65.5	68.8
-0.5	99.6	4.8	96.3	97	6	98.7	98.6	79.1	3.9	33.7	38.8	15.8	58.4	62.7
-0.3	94.9	4.9	72.6	74	6	79.9	79.6	67.5	4.9	28.2	30.1	17.7	46.7	53.1
0.3	97.5	4.7	90.9	92	6	93.8	94.1	50.3	4.9	20	15.1	16.4	37.2	32.7
0.5	99.7	4.8	99.8	99.8	6	99.8	99.8	60.3	4.4	25.3	22.4	13.6	46.3	41.3
0.7	100	4.9	100	100	6.3	100	100	74.8	3	34.5	32.1	11.4	60.5	56.8
$W_{correct}: 10 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 5 \text{ nn}$														
-0.7	99	4.6	97.1	96.3	6.3	98.3	97.7	86.1	4.2	54.7	53.2	13.4	75.5	74.2
-0.5	96.1	4.6	81.1	79	6.3	85.2	84.6	78.1	4.6	40.5	39.4	13.8	63.4	62.6
-0.3	78.7	4.8	38.5	37	6.1	44.5	43.2	62.2	4.6	20.6	21.5	14.7	43.3	42.9
0.3	87	4.7	61.7	61	6.2	67.6	67.5	40.4	3.7	10.4	10.8	15.9	23.7	25.3
0.5	98.1	4.8	94.8	93.9	6.4	95.4	95.7	54.3	3.6	22.2	22.4	13.8	36.8	39.3
0.7	99.7	4.9	99.5	99.6	6.9	99.5	99.6	71.4	3.5	38.1	41.4	12.1	54.9	60
$W_{correct}: 10 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 5 \text{ nn}$														
-0.7	99.4	4.4	97.2	96.7	7	98	97.6	88.1	3.1	53.8	51.6	11.2	75.8	73.6
-0.5	96.1	4.6	81	78.3	7	85.2	84.1	79.6	3.5	40	38	12	63.5	60.7

Continued on next page

Table 3 – Continued from previous page

$\lambda$	First Predictor						Second Predictor							
	% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)			% Corr.	Boot. Rej. Rte (in %)			As. Rej. Rte (in %)		
		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$		$H_0: M_1$	$H_0: M_2$	$H_0: M_3$	$H_0: M_1$	$H_0: M_2$	$H_0: M_3$
-0.3	77.4	4.5	38.7	35.1	6.9	45.1	42.2	64.2	3.9	20.7	20.1	13	43.7	42.2
0.3	86.9	5	62.2	61.2	6.7	68.6	67	41	4.2	9.8	10.4	13.9	23.3	23.6
0.5	98.4	5	94.4	94.3	6.6	95.4	95.6	56.4	3.7	22	22.4	11.3	37.6	37.7
0.7	99.7	5.2	99.5	99.6	6.7	99.6	100	72.7	3	38.4	39.4	9	55.8	58.2
$W_{correct}: 10 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 5 \text{ nn}$														
-0.7	99.2	4.9	97	96.6	6	97.8	97.9	86.4	3	53.5	51.2	14.9	74.8	74.6
-0.5	96.1	5	80.7	81.4	6	85.1	85.5	76.8	3.8	39.7	37	15.5	63	63.9
-0.3	76.8	5	38.1	37.1	6	44.8	43.5	61.9	4.8	20.5	20.4	16.7	43	43.7
0.3	86.4	4.9	62.6	62.7	6.4	68.5	67	41.5	4.3	10.7	10	17	22.6	25.9
0.5	98.2	5.1	94.4	95	6.5	95.3	95.6	56.2	3.4	21.9	21.9	14	36.9	39.8
0.7	99.7	5.1	99.3	99.5	6.5	99.4	99.7	73.7	2.1	39.3	38.2	11.2	55.6	59.2
$W_{correct}: 10 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 10 \text{ nn}$														
-0.7	99.3	5.5	96.6	97.2	7.9	98.1	98.1	86	4.6	52.1	54.2	14.7	74.4	75.5
-0.5	96.5	5.5	80.3	79.4	7.9	84.4	85.1	78.1	5	39.8	41.6	15.9	63.1	63.8
-0.3	79.8	5.2	38.4	37.1	7.8	45.2	44	63	5.8	22.1	23.3	17.3	44.3	45.1
0.3	85.6	5.2	62.6	62.4	7.9	68.5	68.6	43.8	5.8	13	12.1	16.3	26.7	27.1
0.5	98.2	5.2	94.9	94.7	7.5	96.2	96.9	57.2	4.7	27.3	24.3	15.3	44.7	42.6
0.7	99.4	4.8	99.1	99.3	8.1	99.1	99.3	75.8	3.9	46	43.7	13.6	65	61.4
$W_{correct}: 10 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 10 \text{ nn}$														
-0.7	99.6	5.1	97	97.4	7.2	98.4	98.3	87.7	3.6	51.9	52.4	13.1	75.1	73.8
-0.5	96.6	5.1	80.1	78.9	7.2	84.4	83.7	80.3	4.1	39.8	39.9	13.9	63.7	62.1
-0.3	78.1	5.1	38.8	37.2	7.2	45.2	44	63.8	4.7	20.6	23	14.9	44.5	43.6
0.3	86.2	5	62.7	61.3	6.9	68.4	67.8	44.5	4.7	13.2	12.2	14.8	26.5	24.8
0.5	98.4	5.1	94.8	95.3	6.9	96.2	96.4	60.1	4	27.2	24.6	13	44.8	39.9
0.7	99.4	5	99.2	99.4	7.5	99.1	99.4	77	3.4	46	42.2	11.6	65.8	61.4
$W_{correct}: 10 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 10 \text{ nn}$														
-0.7	99.4	5.6	96.9	97.3	8.4	98.3	98.3	85.4	3.3	52.5	50.5	16.1	74.7	75.2
-0.5	97	5.5	80	82.8	8.4	84.6	86	76.6	4	40.1	38.7	16.6	62.9	64.6
-0.3	77.2	5.6	38.9	38.4	8.5	45.1	44.8	61.2	4.7	21.3	21.3	17.5	44.9	45.6
0.3	85.7	5.9	62.5	63.4	8.5	69.3	67.9	43.2	4.9	13.2	10.9	17.3	27.3	27.9
0.5	98.7	5.8	94.8	95.7	8.4	96.3	96.2	58.6	4.4	27.4	23.5	14.4	45.1	43.7
0.7	99.5	5.9	99.2	99.5	8.5	99.2	99.5	77.2	2.2	47.1	41.9	12	65.4	62.9
$W_{correct}: 10 \text{ nn}; W_{2n}: 5 \text{ nn}; W_{3n}: 20 \text{ nn}$														
-0.7	99.3	4.9	96.9	96.6	6.9	98.4	98.2	86.1	4.7	51	53.5	15.4	73.8	73.9
-0.5	96.8	5.1	78.5	80	6.8	83.6	84.7	77.6	5.5	39	42.8	16.3	61	63.3
-0.3	79.3	5	39	37.8	6.8	44.9	44.8	64	5.9	22.4	24.4	17.4	44	45.5
0.3	86.8	4.7	62.7	61.7	7	68.7	67.9	45.8	7	14.7	12.9	15.7	30.8	29.3
0.5	98.4	4.8	95.1	94.7	7.1	95.9	96.7	56.5	6.3	26.7	25.8	15.3	44.6	44.1
0.7	99.8	4.9	99.5	99.7	7.5	99.6	99.7	74.2	4.4	41.7	43.5	13.7	60.4	61.4
$W_{correct}: 10 \text{ nn}; W_{2n}: 10 \text{ nn}; W_{3n}: 20 \text{ nn}$														
-0.7	99.5	5.3	97.5	97	7.6	98.3	98.2	87.3	4.7	50.3	52.1	12.9	73.5	72.6
-0.5	96.8	5.3	79.4	78.9	7.5	84.4	83.6	80.2	4.9	39	40.8	13.8	61.6	62.7
-0.3	78.9	5.3	39.2	39	7.5	46	44.9	63.7	5.2	22.3	23.4	15.4	44.3	43
0.3	86.5	5.3	63.2	61.3	6.9	68.5	66.9	45.9	5.9	13.9	12.5	15.9	31.4	27.8
0.5	98.8	5.3	95.2	95.6	7.1	96.3	96.5	60.3	4.9	26.5	25.1	14.8	45.3	42.4
0.7	99.8	5.3	99.6	99.7	7.2	99.7	99.7	74.9	4.1	42.1	42	12.6	61.4	62.3
$W_{correct}: 10 \text{ nn}; W_{2n}: 20 \text{ nn}; W_{3n}: 20 \text{ nn}$														
-0.7	99.7	4.4	97.4	97.6	6.3	98.4	98.3	84.8	3.7	51.3	51.3	15.5	72.9	74.6
-0.5	97.3	4.5	79.4	82.5	6.3	85	85.8	75.5	5.1	39.5	39.4	17.2	61.8	63.3
-0.3	77.9	4.5	40	40.2	6.4	44.9	45.2	62.4	5.6	21.8	23.3	18.6	43.5	45.1
0.3	87	4.7	63.6	63.5	6.1	68.8	67.2	43.9	5.1	13.9	12.5	17.9	31.6	28.5
0.5	98.9	4.8	95.4	95.6	6	96.3	96.4	57.7	4.4	26.7	24.1	15.4	45.1	44
0.7	99.7	4.9	99.5	99.7	6.2	99.5	99.7	74.5	2.8	41.8	41.9	13.4	61	61.2

Tables 2 and 3 present the performance of the  $J$  and  $MJ$  tests when constructed from the first or the second predictor, for each of the two sample sizes. These tables report both the results based on bootstrap

inference (Boot Rej. Rte) and asymptotic inference (As. Rej. Rte). Bootstrapped inference is based on 399 replications, as suggested in Davidson and MacKinnon (2002) for the case of Monte Carlo experiments.

We first analyze the results of Table 2, corresponding to  $n = 60$ . For the first predictor, the bootstrapped size ( $J$  test associated to  $M_1$ ) is always around the theoretical value of 5% (between 4 and 5.8%), no matter the value of the cross-sectional dependence parameter  $\lambda$ , nor the combination of interaction matrices used to design the three SAR models. As expected, the bootstrapped rejection rates of models 2 and 3 (representing a measure of the power of the  $J$  test) are very high for high spatial dependence (positive or negative) but decrease as  $\lambda$  goes to 0 in absolute value. Indeed, for smaller value of the parameter, it is harder to distinguish between alternative connectivity matrices as cross-sectional dependence is weaker. The high power is especially observed when the correct interaction matrix is sparse ( $W_{correct} = 5nn$ ) but decreases as the correct matrix gets denser ( $W_{correct} = 10nn$ ). The size based on asymptotic inference is always liberal, ranging between 8.1 and 10%, a result confirming those obtained by Godfrey (1998); Fan and Li (1995) and Davidson and MacKinnon (2002). Finally, even though rejection rates for models 2 and 3 are slightly higher than for the bootstrapped inference, they cannot be directly compared.<sup>17</sup>

Regarding the second predictor, we observe an important size-distortion for the  $J$  test based on asymptotic distribution. Indeed, the size varies between 11 and 27.7%, thus limiting the use of this statistic when based on this predictor. By contrast, the bootstrapped size is much closer to the theoretical 5%, even though its behavior is more volatile (i.e. ranging between 2 and 6.4 %) than for the first predictor. Further, rejection rates of models 2 and 3 are much lower for the second predictor, both for bootstrapped and asymptotic based inference.

Turning to Table 3, which summarizes the results for  $n = 100$ , we observe a lower discrepancy between bootstrapped and asymptotic based sizes for the first predictor. This leads to smaller power differences between the two approaches. We also note higher rejection rates for models 2 and 3 for small cross-sectional dependence compared to  $n = 60$ . Further, the power difference considering  $W_{correct} = 5nn$  and  $W_{correct} = 10nn$  is not present anymore, indicating that it is the relative sparseness of the correct matrix that might matter, rather than the absolute one.

Regarding the second predictor, the bootstrapped size is closer to the theoretical 5% than its asymptotic

---

<sup>17</sup> To compare the power of tests, the sizes must be held constant (see Davidson and MacKinnon, 2002, p.188).

counterpart, but it fluctuates more than that of the first predictor. Further, the power of the test is much lower, no matter the value of the cross-sectional dependence parameter  $\lambda$  nor the set of interaction matrices considered.

Lastly, Tables 2 and 3 assess the performance of the  $MJ$  test to select the right model. The variable  $\%Corr.$  reports the percentage of times the  $MJ$  test picks up the right model. Focusing on the first predictor and on Table 2 ( $n = 60$ ), we note a very good performance of the  $MJ$  test to select the right SAR model. This is especially true for moderate ( $|\lambda| = 0.5$ ) and high cross-sectional dependence ( $|\lambda| = 0.7$ ), where the percentage of correct selection is always above 90% when the correct interaction matrix is sparse ( $W_{correct} = 5nn$ ). When the correct matrix is denser ( $W_{correct} = 10nn$ , implying that each point is neighbor with around 17% of the sample), the percentage of correct model selection decreases, nevertheless remaining at decent levels. However, this result is peculiar to small sample sizes. Looking at Table 3, we do not notice any significant differences between  $W_{correct} = 5nn$  and  $W_{correct} = 10nn$ , and observe a percentage of correct selection of at least 95 %. This strengthens our earlier point about the importance of the relative sparseness rather than its absolute counterpart. Looking at low cross-sectional dependence ( $|\lambda| = 0.3$ ) we note a lower percentage of correct selection, but remaining very good compared to the power of the  $J$  test. For instance, for the case where  $W_{correct} = 5nn$ ,  $W_{2n} = 10nn$  and  $W_{3n} = 10nn$  and  $\lambda = -0.3$ , the  $MJ$  test selects the right model 73.7% of the time while the rejection rate for models 2 and 3 are of 30.3% and 34.6% respectively. This indicates that without the  $MJ$  test, we would have faced a model selection problem in nearly 70% of the time.

To summarize, the  $J$  test constructed from the first predictor performs better than the one based on the second predictor, both in terms of size and power. Further, we observe a better behavior of the bootstrapped size compared to asymptotic-based size. Results also seem to indicate that the relative sparsity of the correct matrix affects the results, rather than its absolute sparsity. As such, for  $n = 60$ , a denser correct matrix (constructed from the 10 nearest neighbors) has adverse effect on both the power of the  $J$  test and on the performance of the  $MJ$  test, but this effect disappears when the sample size increases. Lastly, the  $MJ$  test performs very well in identifying the correct SAR model, even for small sample sizes.



## 6. Application to the generalized Schumpeterian growth model

The purpose of this section is to apply our testing procedure to the multi-country Schumpeterian growth model with worldwide interactions, developed by Ertur and Koch (2011). These authors propose an integrated theoretical and methodological framework characterized by technological interactions to explain growth processes from a Schumpeterian perspective. Their generalized Schumpeterian multi-country growth model, extending Howitt (2000), explicitly accounts for global technological interdependence implied by R&D spillovers. Moreover they show that a multi-country Solow growth model with technological interdependence similar to that proposed by Ertur and Koch (2007) is actually nested in the generalized model leading to a simple test to choose between the two. More specifically, they show that when the R&D expenditures have no effect on the growth rate of technology, the multi-country Schumpeterian growth model reduces to the multi-country Solow growth model.<sup>18</sup>

Their empirical results show that the multi-country pure Solow growth model is rejected in favor of the multi-country Schumpeterian growth model, once both models are integrated in a unified theoretical and methodological framework characterized by technological interdependence. They finally estimate the TFP equation implied by their multi-country Schumpeterian growth model as well as the international R&D spillovers implied by technological interdependence using two different interaction matrices based on geographical distance and trade flows. The main result they underline is the presence of multiple local technological leaders: besides USA, Germany and Japan are also playing that role in their respective area of influence.

### 6.1. The econometric specification

Due to space limitation, we focus on the TFP equation implied by the multi-country growth model (see Ertur and Koch, 2011, eq. 59, p.234):

$$\ln TFP_i = \beta_0 + \beta_1 \ln \frac{s_{K,i}}{n_i + 0.05} + \beta_2 \ln s_{A,i} + \beta_3 \ln n_i + \gamma H_i \sum_{j \neq i}^n v_{ij} \ln TFP_j + \varepsilon_i \quad (20)$$

---

<sup>18</sup>The interested reader can refer to Ertur and Koch (2011) for the full development of the multi-country Schumpeterian growth model.

where  $\ln TFP_i$  is the logarithm of Total Factor Productivity of country  $i$  at steady state (taking one third as the capital's share in income);  $\beta_1$  is the coefficients associated with the logarithm of the investment rate divided by the effective depreciation rate  $\ln \frac{s_{K,i}}{n_i+0.05}$ ,  $\beta_2$  is the coefficient associated with the logarithm of the investment rate in the research sector  $\ln s_{A,i}$  and  $\beta_3$  is the coefficient associated to the logarithm of the working-age population growth rate  $n_i$ .  $H_i$  is the stock of human capital, designed to measure the capacity of absorption of a new technology by country  $i$ . Finally  $\gamma$  is the autoregressive interaction parameter. In matrix form, the model, in cross-section, is written as follows:

$$y = \beta_0 \iota + \beta_1 S_K + \beta_2 S_A + \beta_3 N + \gamma W y + \varepsilon \quad (21)$$

where  $y$  is the  $(n \times 1)$  vector of the logarithms of  $TFP$ ;  $\iota$  is the unit vector,  $S_K$  is the vector of the logarithms of the investment rates in physical capital divided by the effective depreciation rates,  $S_A$  is the vector of the logarithms of expenditures in the research sector and  $N$  is the vector of the logarithms of working-age population growth rates.  $W$  is the interaction matrix modeling technological interactions as a function of the new technology absorption capacity of the receiving country and some ad hoc measure of similarity between countries, with general term  $w_{ij} = H_i v_{ij}$ . In this application, We consider different definitions of similarity  $v_{ij}$  and select the most relevant one given the information contained in the sample.

## 6.2. Data and interaction matrices

For the sake of comparability we use the same database as Ertur and Koch (2011). The data are extracted from the Heston et al. (2006) Penn World Tables (PWT version 6.2), containing information on real income, investment and population for a large number of countries. Data on R&D expenditures come from the World Investment Report (2005) of the United Nations Conference on Trade and Development (UNCTD). The sample contains 58 countries and includes 21 OECD countries over the period 1990-2003.<sup>19</sup> The explanatory variables are constructed as follows. To measure  $n_i$ , for  $i = 1, \dots, n$ , as the average growth of the working-age population (ages 15 to 64), it is necessary to compute the number of workers as:  $RGDPCH \times POP/RGDPW$ , where  $RGDPCH$  is real GDP per capita computed by the chain method,  $RGDPW$  is real-chain GDP per worker, and  $POP$  is the total population. The saving rate  $s_{K,i}$ , for

---

<sup>19</sup>Due to missing data for the computation of the linguistic interaction matrix, we exclude South Korea from the sample as Korean is generally admitted to be a language isolate.

$i = 1, \dots, n$ , is measured as the average share of gross investment in GDP over the period as in Mankiw et al. (1992). The variable  $s_{A,i}$ , is measured as the average share gross domestic expenditure on R&D (GERD) relative to GDP over the 1991-2001 period. Finally, the human capital stock is measured using the Mincerian equation, as proposed by Hall and Jones (1999) or Caselli (2005). We use its value measured in 1990 for the population between 25 and 64 in each country, provided by the database developed by Cohen and Soto (2007) and updated by Cohen and Leker (2014).<sup>20</sup> The dependent variable is defined in Ertur and Koch (p.234, 2011) and computed as  $\ln TFP_i = \ln y_i - 0.5 \ln \frac{s_{K,i}}{n_i + 0.05}$  where  $y_i$  is the real income per worker in 2003. Table 4 provides some descriptive statistics over the variables used in our specification.

Table 4: Descriptive statistics					
	Min	Max	Mean	Median	St. err.
$\ln TFP$	3.2444	4.5469	4.1123	4.1789	0.3223
$H_i$	0.2389	13.3311	7.6254	7.2303	3.1882
$\ln s_A$	-3.6815	-1.3694	-2.5357	-2.6789	0.6779
$\ln n$	-1.5016	-1.3463	-1.4333	-1.4397	0.0427
$\ln \frac{s_K}{n+0.05}$	-0.1363	0.3867	0.2169	0.2284	0.1229

The first interaction matrix we consider ( $W1$ ) is based on genealogical or "genetic" distance following Spolaore and Wacziarg (2009, 2013).<sup>21</sup> This notion of distance corresponds to the time elapsed since two populations shared a common ancestor. In other words, the genetic distance is a measure of relatedness between populations over time. It can be interpreted as a general metric for average differences in characteristics transmitted across generations. Moreover they argue that: "...by its very definition, genetic distance is an excellent *summary statistic* capturing divergence in the whole set of implicit beliefs, customs, habits, biases, conventions, etc. that are transmitted across generations - biologically and/or culturally - with high persistence." (Spolaore and Wacziarg, 2009, p.471). The general idea is that more closely related societies are more likely to learn from each other and adopt each other's innovations. Spolaore and Wacziarg (2009, 2013) show that this general genealogical measure between populations may then explain income differences today. Their interpretation is that the genetic distance captures the barriers to the diffusion of development.

Several measures of genetic distance have been proposed in the literature related to population genetics.

---

<sup>20</sup>This database is available at: <http://www.parisschoolofeconomics.eu/fr/cohen-daniel/base-de-donnees-internationale-education/>.

<sup>21</sup>The term genetic distance comes from the literature related to population genetics and is also used in linguistics to refer to a measure of relatedness.

The most popular are FST distance or “co-ancestor coefficient” (Wright, 1965) and Nei’s distance (Nei, 1973). These two measures have slightly different theoretical properties but their correlation is very strong. Thus, the choice of one these should not affect the results. The basic unit of these measures is the allele, which is a form taken by a gene. Differences in allele frequencies hence form the basis of summary measures of calculating distance between populations. The genetic distance matrix  $W1$  used here is based on the differentiation coefficient  $G_{ST}$  of Nei which corresponds to the probability that two randomly selected variants in a population are different. Nei (1973) propose indexes based on allele frequencies that describe the level and organization of diversity within populations. These indexes portray the total diversity ( $h_T$ ), the average intra-population diversity ( $h_S$ ) or differentiation between populations ( $G_{ST}$ ).  $G_{ST}$  takes a value equal to zero if and only if the distributions of alleles are identical between the two populations, whereas it is positive when the distributions of alleles are different. A greater distance measure  $G_{ST}^W$  reflects a longer separation between populations. However, many countries, such as United States and Australia, are composed of sub-populations that are genealogically distant. It is thus preferable to use a weighted measure of genealogical distance  $G_{ST}$ . The interaction matrix  $W1$  is finally constructed as a decreasing function of the weighted genealogical distance  $G_{ST}^W$  between countries whose functional form is the negative exponential of the distance. The general term of this matrix  $W1$  is defined as  $w1_{ij} = H_i v1_{ij}$  with:

$$v1_{ij} = \begin{cases} 0 & \text{if } i = j \\ e^{-g_{ij}} / \sum_{j \neq i} e^{-g_{ij}} & \text{otherwise} \end{cases} \quad (22)$$

where  $g_{ij}$  represents the weighted genetic distance  $G_{ST}^W$  between countries  $i$  and  $j$  and  $H_i$  is the human capital stock of the receiving country  $i$ .

The second interaction matrix we consider ( $W2$ ) is based on a linguistic proximity concept following Melitz and Toubal (2014). They summarize the evidence about the linguistic influences in an aggregate index, named Common Language index (CL), resting strictly on exogenous linguistic factors. This summary index (CL) is obtained using three indexes: Common Official Language (COL), Common Native Language (CNL) and Language Proximity (LP).<sup>22</sup> The Common Language index (CL) ranges between 0 and 1. More precisely, CL takes a value equal to zero if the degree of linguistic proximity between two

---

<sup>22</sup>Data on Common Language are available at [http://www.cepii.fr/cepii/en/bdd\\_modele/presentation.asp?id=19](http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=19).

countries is null, whereas CL tends to 1 when the linguistic proximity between these two countries is high. The general term of this matrix  $W2$  is defined as  $w2_{ij} = H_i v2_{ij}$  with:

$$v2_{ij} = \begin{cases} 0 & \text{if } i = j \\ CL_{ij} / \sum_{j \neq i} CL_{ij} & \text{otherwise} \end{cases} \quad (23)$$

where  $CL_{ij}$  represents the Common Language Index (CL) between countries  $i$  and  $j$  and  $H_i$  is the human capital stock of the receiving country  $i$ .

The last considered interaction matrix  $W3$  is based on trade flows. Indeed, trade flows may proxy multi-country technological interactions.<sup>23</sup> The general term of this matrix is defined as  $w3_{ij} = H_i v3_{ij}$  where:

$$v3_{ij} = \begin{cases} 0 & \text{if } i = j \\ m_{ij} / \sum_{j \neq i} m_{ij} & \text{otherwise} \end{cases} \quad (24)$$

with  $m_{ij}$  defined as the average imports of country  $i$  coming from country  $j$  and computed over the 1990-2000 period to prevent endogeneity problems that might arise, using the bilateral trade flows database from CEPII. Ertur and Koch (2011) use a different international trade database and define a block triangular interaction matrix assuming that the North club (OECD countries) diffuses its knowledge to the South club (the rest of the countries of their sample), but the opposite does not occur (see Appendix 2, p.250, Ertur and Koch (2011)). Here we relax this assumption and further allow the South club to diffuse its knowledge to the North club. Table 5 reports the correlation measure between spatial lags constructed from the product of each interaction matrix with a standard normal variable  $u$ , as suggested by LeSage and Pace (2014). We observe that the 3 considered matrices are not highly correlated, meaning that they represent different interaction schemes.

---

<sup>23</sup> The choice of this matrix is justified by studies of Grossman and Helpman (1991); Coe and Helpman (1995) and Coe et al. (1997) which show that international trade may be seen as a major diffusion vector of technological progress.

Table 5: Correlation measure between spatial lags

	$W_1u$	$W_2u$	$W_3u$
$W_1u$	1.000	-	-
$W_2u$	0.174	1.000	-
$W_3u$	0.167	0.095	1.000

### 6.3. Econometric results

The TFP equation (20) implied by the multi-country Schumpeterian growth model is estimated by the RGMM procedure described in section 2 using the interaction matrices based on genealogical distance ( $W_1$ ), linguistic distance ( $W_2$ ), and trade flows ( $W_3$ ). Each matrix has been normalized by the minimum between the maximal values of the row sums and column sums, as suggested in Kelejian and Prucha (2010, Lemma 2, p.56). Estimation results are presented in Table 6. They are consistent with those obtained by Ertur and Koch (2011) using different interaction matrices, underlying the robustness of the model.<sup>24</sup> The coefficient associated with R&D expenditures is significant whatever the interaction matrix considered. Its estimated value ranges from 0.159 to 0.170. We also note that the spatial autocorrelation parameter is significant and has an estimated value ranging from 0.151 to 0.169. Thus, the TFP of one country cannot be considered as independent from that of other countries. The coefficients of the investment rate divided by the effective depreciation rate and of the working-age population growth rate are non-significant as in Ertur and Koch (2011, Table 2, p.242).

In order to provide considerable help in the selection of the interaction matrix, we apply the  $J$  tests to equation (20) and then use the  $MJ$  procedure to select the most adequate interaction matrix (out of the three considered). The results are presented in Table 7. We first observe that bootstrapped p-values are much higher than p-values based on asymptotic distribution ( $\chi^2_2$ ). This result is in line with what we found in the Monte Carlo experiments. Further, none of the three  $J$  tests are significant, implying a decision problem with respect to the selection of the interaction matrix. To overcome this decision problem, we use the  $MJ$  test and select the model for which the associated  $J$  statistic has the minimum value. This minimum value (in bold) is obtained for the model with  $W_3$ , namely the interaction matrix based on a function of the capacity to absorb new technologies (measured by the stock of human capital of the

---

<sup>24</sup>Note that Ertur and Koch (2011) estimate also a constrained model where they introduce some structural theoretical restrictions  $\beta_1 = \beta_2 = \beta_3 = \beta$  (p.241), which would take a lot of space to explain here. Therefore due to space limitation and for more simplicity, we only estimate here the unconstrained model. The interested reader could refer to this paper. Note also that the non-significance of  $\beta_1$  and  $\beta_3$  is a shared feature between the two papers.

Table 6: Estimation Results

Estimation method Interaction matrix	SAR-RGMM		
	(W1)	(W2)	(W3)
Constant	4.615 (3.220)	4.776 (3.367)	4.638 (3.247)
$\ln(s_K/(n + 0.05))$	-0.055 (-0.142)	-0.033 (-0.087)	-0.089 (-0.228)
$\ln s_A$	0.159 (2.068)	0.170 (2.208)	0.165 (2.142)
$\ln n$	0.379 (0.323)	0.452 (0.487)	0.355 (0.379)
$Wy$	0.151 (3.274)	0.151 (3.311)	0.169 (3.273)
$n$	58	58	58
Pseudo- $R^2$	0.640	0.635	0.635

t-stats between brackets. Pseudo- $R^2$  is the squared correlation between predicted and actual values.

receiving country) and on trade flows.

Table 7:  $MJ$  tests

$H_0$ :	W1	W2	W3
$H_A$ :	W2,W3	W1,W3	W1,W2
J-test	3.048 (0.380) (0.218)	2.906 (0.427) [0.234]	<b>2.458</b> (0.546) [0.293]

Asymptotically-based p-values between square brackets and Bootstrapped p-values, using 999 replications are shown between brackets.

On the basis of econometric results in Tables 6 and 7 and equation (3), we compute the impacts of home and foreign R&D expenditures on the TFP of a given country using  $W3$ , which is favored by the  $MJ$  test. We then obtain the  $n \times n$  estimated matrix of elasticities, using the estimated coefficients of the econometric model. Results for OECD countries plus Brazil and China, displaying significant impacts (at the 5% level) are presented in Table 9.<sup>25</sup>

Table 9 is divided in two parts: the upper part displays R&D spillovers within the club of OECD countries (North-North R&D spillovers) while the lower part shows R&D spillovers from OECD countries

<sup>25</sup>Inference on impacts is based on simulating 1000 values of these impacts. As the estimators are shown to be normally distributed, for these simulations, we use a normal distribution centered on the estimated values of the parameters and with the covariance matrix of the estimated parameters.

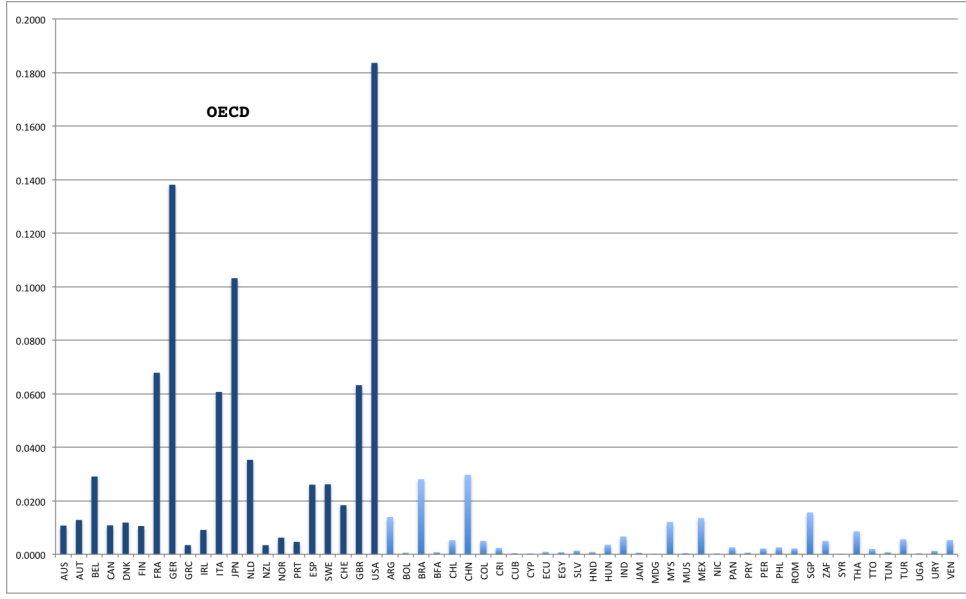


Figure 1: Indirect impacts of a change of R&D expenditures on the TFP of all countries using the trade based interaction matrix ( $W3$ )

to countries not belonging to this club (North-South R&D spillovers). Results for China and Brazil are added in the last columns to give an insight about R&D spillovers from non OECD emerging countries to countries belonging to the North club (South-North R&D spillovers) and to the South club (South-South R&D spillovers). The highlighted (in bold) array elements on the main diagonal correspond to the direct elasticities, including feedback effects, of the TFP of a given country with respect to its own R&D expenditures.

We first note that the direct elasticities are all significant and only slightly higher than  $\hat{\beta}_2 = 0.1653$  when we take into account cross-sectional dependence, because of higher order feedback effects in the expansion (4). Even though these direct effects are heterogeneous by nature, in this application we do not observe much difference between countries, due to the fact that feedback effect are of small scale compared to the estimated value of  $\hat{\beta}_2$ . It is however interesting to note that those feedback effects are the highest for USA, Germany, Japan, Canada, France and United Kingdom (see Table 9). For OECD countries, they range from 0.1653% for Greece to 0.1666% for USA, which benefits of the highest feedback effects, while the average direct impact including feedback effects is 0.1656%, i.e. on average, a 1% increase of their R&D expenditures yields a 0.1656% increase of TFP. For countries not belonging to this club, on average, a 1% increase of their R&D expenditures yields a 0.1654% increase in TFP.

Figure 1 presents results for significant indirect impacts, for the whole sample, which are our main



concern here. Using Table 9, the flow of knowledge between countries  $i$  and  $j$  goes from the country in column  $j$ , the emitting country, to the country in row  $i$ , the receiving country.

Second, with respect to indirect elasticities, i.e. the off-diagonal terms of the matrix in Table 9, we note that the United States is the country, among OECD countries, which diffuses the most of its R&D to other countries, followed by Germany and Japan (Figure 1). The indirect impact of a 1% increase of R&D expenditures in USA on the TFP of all other countries is 0.1836%, whereas this impact is 0.1381% for Germany and 0.1032% for Japan (see sums of the corresponding columns of Table 9 and Figure 1). Among the OECD group, countries which diffuse the least of their R&D are Greece, New-Zealand and Portugal. Note also that the share of total intra OECD impact in the total impact ranges from 68% for USA to 99% for Portugal. For non-OECD countries, it is interesting to note that for China, the indirect impact of a 1% increase of R&D expenditures on the TFP of all other countries is 0.0297%, an impact that is higher than impacts reported for most of the OECD countries. This is also the case for Brazil (0.0281%).

In addition, a 1% increase in R&D expenditures in USA yields a 0.0180% increase in Canadian TFP, whereas a 1% increase in R&D expenditures in Canada only yields a 0.0054% increase in US TFP. This asymmetric effect is mainly due to the asymmetric nature of trade between USA and Canada. On average, for the period under study, only 19.82% of the total imports of USA come from Canada, whereas 67.27% of the total imports of Canada come from USA. Note also that the capacity of absorption of new technologies, as measured by human capital stocks, is slightly higher in USA (12.75) than in Canada (12.44). Figure 2 displays this capacity of absorption for all countries in the sample. Even though the value and significance of estimated coefficients in Table 6 are robust to the choice of interaction matrix, it is nevertheless highly important to select the right interaction matrix. To make our point, we show in Figures 3 and 4 the significant indirect effects obtained for the two other considered interaction matrices. We note that their patterns are completely different from each other, implying different policy recommendations. Further, the three local technological leaders identified using the interaction matrix based on trade do not appear anymore for interaction matrices based on linguistic or genealogical similarity.

It is also possible to interpret the results in terms of clubs. Inside the North club (OECD countries), the average indirect impact is 0.023, i.e. on average, a 1% increase of the R&D expenditures of a OECD country yields a 0.023% increase of the TFP of all other OECD countries. Inside the South club (non-OECD countries), on average, a 1% increase of the R&D expenditures yields a much lower 0.004% increase

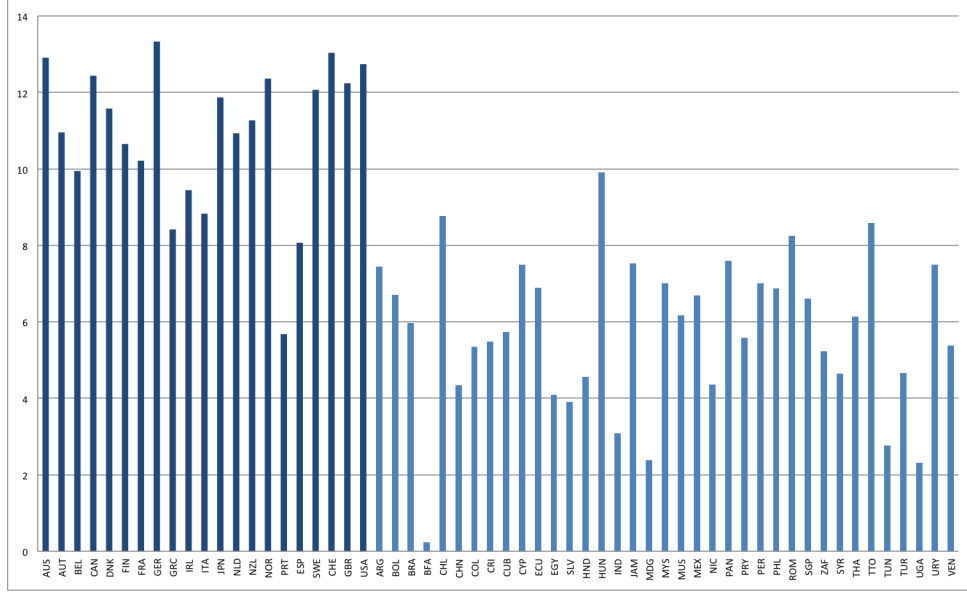


Figure 2: Human capital stock ( $h_i$ )

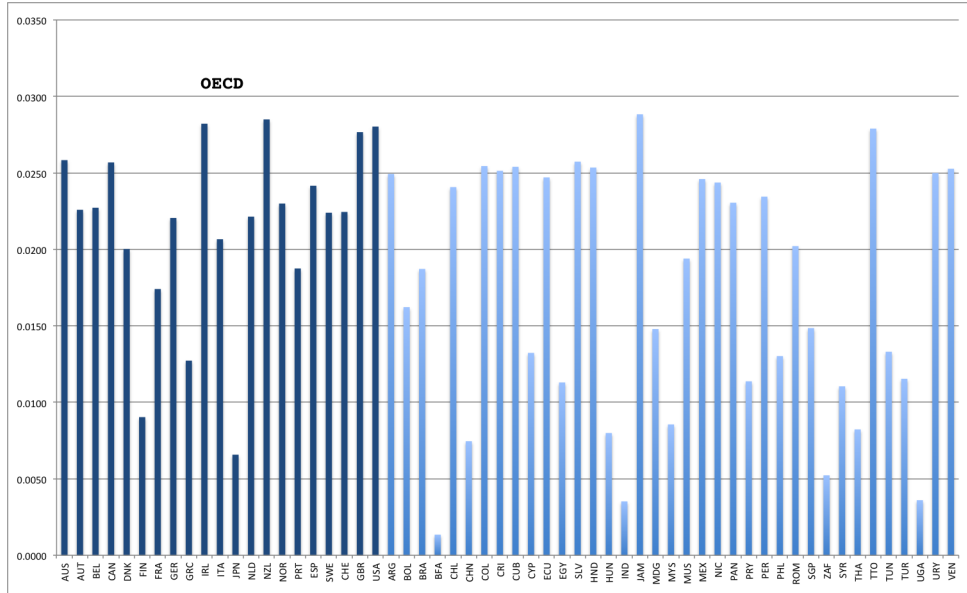


Figure 3: Indirect impacts of a change of R&D expenditures on the TFP of all countries using the linguistic distance based interaction matrix ( $W2$ ).

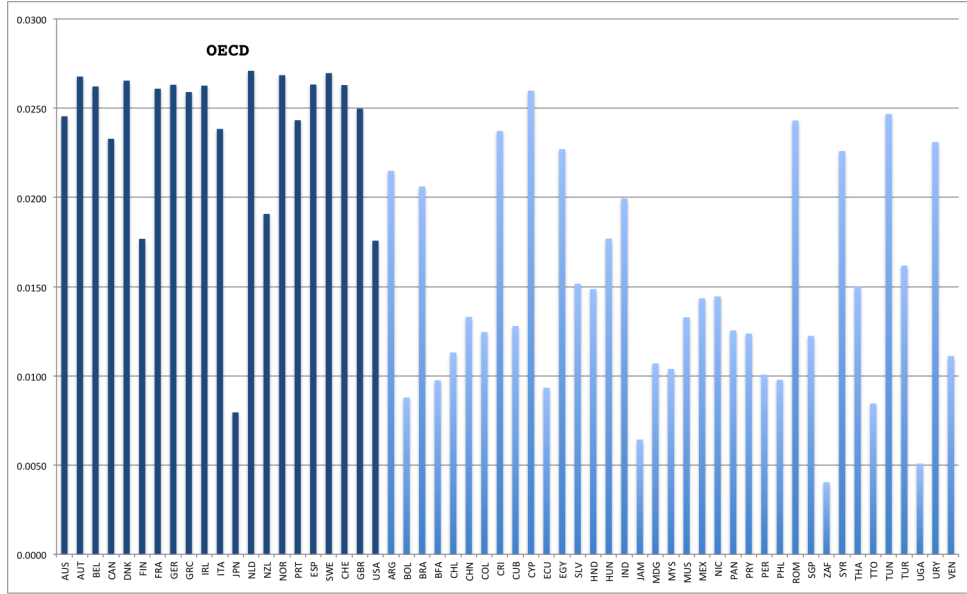


Figure 4: Indirect impacts of a change of R&D expenditures on the TFP of all countries using the genealogical distance based interaction matrix ( $W1$ ).

of the TFP of the other countries. Not surprisingly, the North-South average indirect impact is on average 0.059%, much higher than the South-North average indirect impact, which is 0.002% (see Table 8). Again, not surprisingly this North-South club heterogeneity comes from indirect effects, the direct effects being pretty similar across clubs.

Table 8: Average significant impacts for Clubs

Club	Mean	Median
<b>North - North</b>		
Direct impact	0.166	0.165
Indirect impact	0.023	0.012
Total impact	0.188	0.178
<b>North - South</b>		
Indirect impact	0.059	0.059
<b>South - South</b>		
Direct impact	0.165	0.165
Indirect impact	0.004	0.001
Total impact	0.169	0.167
<b>South - North</b>		
Indirect impact	0.002	0.001

Third, those results are essentially due to the weight of the United States in the international trade pattern. The United States R&D diffusion impacts are higher for countries belonging to Central and South-America, such as Canada, Mexico, Trinidad and Tobago, Jamaica, Costa-Rica and Honduras for

instance than for other countries belonging to the South Club.<sup>26</sup> We also note the role played by the human capital stock in enhancing the absorption capacity of international R&D diffusion since the impact on Canada (0.018%) is higher than the impact on Mexico (0.011%) although the latter has an higher import share from the United States. These results for United States show the influence of this country in the American continent, as also underlined by Coe et al. (1997). The elasticities from Japan to South East Asian countries are also higher than the elasticities from Japan to other countries belonging to the South club. The elasticities from Germany to other European countries are higher than the elasticity from Germany to other countries belonging to the south club. These results suggest that the United States is a natural technological leader for Central and Southern American countries, that Japan is the technological leader in South East Asia and that Germany is the technological leader in Europe.

## 7. Conclusions

We propose a solution to the interaction matrix selection problem in spatial econometric models using the RGMM estimation method proposed by Lin and Lee (2010), which consists in a new, robust against unknown heteroskedasticity, version of the  $J$  test first proposed by Kelejian (2008) and modified by Kelejian and Piras (2011). Moreover the common model selection problem that arises with non-nested hypotheses tests is overcome using the approach proposed by Hagemann (2012), called the  $MJ$  test.

We then apply this non-nested hypotheses testing procedure to the the multi-country Schumpeterian growth model proposed by Ertur and Koch (2011) to discriminate between three different types of interaction matrices: genealogic distance, linguistic proximity and bilateral trade flows. The estimation results of the TFP equation implied by this model are robust with respect to the choice of the interaction matrix. However using the  $J$  test to select an interaction matrix leads to a model selection problem. To overcome this problem, we use the  $MJ$  test which indicates that the interaction matrix based on a function of the capacity of absorption of new technologies and trade flows is the most appropriate amongst the three interaction matrices we consider.

The obtained estimation results are consistent with those of Ertur and Koch (2011) and Coe et al. (1997) and highlight the heterogeneity of the international diffusion of knowledge. As already noted by Ertur and

---

<sup>26</sup>In our sample, Canada imports almost 67% from the United States, Mexico, 77%; Trinidad and Tobago, 51%; Jamaica, 54%; Honduras, 56% and Costa-Rica, 53%.

Koch (2011), this empirical evidence cannot be captured by the standard Howitt and Aghion (1998) and Howitt (2000) models, which assume that USA are the unique *global* technological leader, whereas the integrated multi-country Schumpeterian growth model allows the emergence of *local* technological leaders. Using the selected trade flows based interaction matrix, we show that indeed USA, but also Germany and Japan actually play the role of *local local* technological leaders.

We believe that the suggested interaction matrix selection procedure will be helpful for applied researchers.

## Acknowledgements

We would like to thank Matthias Pécot and Yvan Stroppa for valuable research assistance. Also, we benefit from the support of the Centre de Calcul Scientifique en région Centre and from University of Lille to perform all Monte Carlo simulations. The authors also acknowledge the financial support of the Region Centre Val de Loire (APR-IA 2014 00094514). We are also grateful to Ingmar Prucha, Lung-Fei Lee, Fei Jin and the participants of the 14<sup>th</sup> international workshop on Spatial econometrics and statistics in Paris, 7<sup>th</sup> Jean Paelinck Seminar in Saragossa, 26<sup>th</sup> EC<sup>2</sup> conference hosted in Edinburgh, the 5<sup>th</sup> Shanghai Econometrics Workshop and the Asian Meeting of the Econometric Society, Kyoto, 2016.

## References

- Akerlof, G. A. (1997), 'Social distance and social decisions', *Econometrica* **65**, 1005–1027.
- Anselin, L. (1988), 'Model validation in spatial econometrics: a review and evaluation of alternative approaches', *International Regional Science Review* **11**, 279–316.
- Anselin, L. and Bera, A. (1998), Spatial dependence in linear regression models with an application to spatial econometrics, in A. Ullah and D. Giles, eds, 'Handbook of Applied Economics Statistics', Marcel Dekker, New York, pp. 237–289.
- Ashraf, Q. and Galor, O. (2013), 'The "out of africa" hypothesis, human genetic diversity, and comparative economic development', *The American Economic Review* **103**(1), 1–46.
- Atkinson, A. (1970), 'A method for discriminating between models.', *Journal of the Royal Statistical Society, Serie B* **32**, 323–353.
- Badinger, H. and Egger, P. (2011), 'Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances', *Papers in Regional Science* **90**, 213–235.
- Burridge, P. (2012), 'Improving the J test in the SARAR model by likelihood-based estimation', *Spatial Economic Analysis* **7**, 75–107.
- Burridge, P. and Fingleton, B. (2010), 'Bootstrap inference in spatial econometrics: the J-test', *Spatial Economic Analysis* **5**, 93–119.
- Caselli, F. (2005), *Accounting for cross-country income differences*, Vol. 1, Elsevier, chapter 9, pp. 679–741.
- Cavalli-Sforza, L. L., Menozzi, P. and Piazza, A. (1994), *The History and Geography of Human Genes*, Princeton University Press.
- Coe, D. T. and Helpman, E. (1995), 'International R & D spillovers', *European Economic Review* **39**, 859–887.
- Coe, D. T., Helpman, E. and Hoffmaister, A. (1997), 'North-South R % D spillovers', *The Economic Journal* **107**, 134–149.

- Cohen, D. and Leker, L. (2014), Health and education: Another look with the proper data, Discussion Paper 9940, Centre for Economic Policy Research.
- Cohen, D. and Soto, M. (2007), ‘Growth and human capital: good data, good results’, *Journal of Economic Growth* **12**, 51–76.
- Comin, D., Easterly, W. and Gong, E. (2010), ‘Was the wealth of nations determined in 1000 BC?’, *American Economic Journal: Macroeconomics* **2**(3), 65–97.  
**URL:** <http://dx.doi.org/10.1257/mac.2.3.65>
- Cox, D. (1961), ‘Tests of separate families of hypotheses’, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* **1**, 105–123.
- Cox, D. R. (1962), ‘Further results on tests of separate families of hypothesis.’, *Journal of the Royal Statistical Society, Serie B* **24**, 406–424.
- Davidson, R. and Flachaire, E. (2008), ‘The wild bootstrap, tamed at last’, *Journal of Econometrics* **146**, 162–169.
- Davidson, R. and MacKinnon, J. G. (1981), ‘Several tests for model specification in the presence of alternative hypotheses’, *Econometrica* **49**, 781–793.
- Davidson, R. and Mackinnon, J. G. (1985), ‘Heteroskedasticity-robust tests in regression directions’, *Annales de L’insee* **59–60**, 183–218.
- Davidson, R. and MacKinnon, J. G. (2002), ‘Bootstrap J tests of nonnested linear regression models’, *Journal of Econometrics* **109**, 167–193.
- Debary, N. and Ertur, C. (2010), ‘Testing for spatial autocorrelation in a fixed effects panel data model’, *Regional Science and Urban Economics* **40**, 453–470.
- Debary, N. and LeSage, J. (2018), ‘Flexible dependence modeling using convex combinations of different types of connectivity structures’, *Regional Science and Urban Economics* **69**, 48–68.
- Elhorst, J. P., Lacombe, D. J. and Piras, G. (2012), ‘On model specification and parameter space definitions in higher order spatial econometric models’, *Regional Science and Urban Economics* **42**, 211–220.

- Ertur, C. and Koch, W. (2007), ‘Growth, technological interdependence and spatial externalities: theory and evidence’, *Journal of Applied Econometrics* **22**, 1033–1062.
- Ertur, C. and Koch, W. (2011), ‘A contribution to the theory and empirics of schumpeterian growth with worldwide interactions’, *Journal of Economic Growth* **16**, 215–255.
- Fan, Y. and Li, Q. (1995), ‘Bootstrapping J-type tests for non-nested regression models’, *Economic Letters* **48**(2), 107–112.
- Flachaire, E. (2005), ‘Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap’, *Computational Statistics and Data Analysis* **49**(2), 361–376.
- Freedman, D. A. (1981), ‘Bootstrapping regression models’, *The Annals of Statistics* **9**(6), 1218–1228.
- Godfrey, L. G. (1983), ‘Testing non-nested models after estimation by instrumental variables or least squares.’, *Econometrica* **51**, 355–366.
- Godfrey, L. G. (1998), ‘Tests of non-nested regression models: some results on small sample behaviour and the bootstrap’, *Journal of Econometrics* **84**, 59–74.
- Godfrey, L. G. and Pesaran, H. M. (1983), ‘Tests of non-nested regression models: Small sample adjustments and monte carlo evidence’, *Journal of Econometrics* **21**, 133–154.
- Grossman, G. M. and Helpman, E. (1991), *Innovation and growth in the global economy*, MIT Press, Cambridge, MA.
- Hagemann, A. (2012), ‘A simple test for regression specification with non-nested alternatives’, *Journal of Econometrics* **166**, 247–254.
- Hall, R. E. and Jones, C. I. (1999), ‘Why do some countries produce so much more output per worker than others?’, *The Quarterly Journal of Economics* **114**, 83–116.
- Hansen, L. P. (1982), ‘Large sample properties of generalized method of moments estimators’, *Econometrica* **50**, 1029–1054.
- Hazır, C. S., LeSage, J. and Autant-Bernard, C. (2018), ‘The role of R&D collaboration networks on regional knowledge creation: Evidence from information and communication technologies’, *Papers in Regional Science* **97**(3), 549–567.



- Heston, A., Summers, R. and Aten, B. (2006), 'Penn world table version 6.2'.
- Howitt, P. (2000), 'Endogenous growth and cross-country income differences', *The American Economic Review* **90**, 829–846.
- Howitt, P. and Aghion, P. (1998), 'Capital accumulation and innovation as complementary factors in long-run growth', *Journal of Economic Growth* **3**, 111–130.
- Jin, F. and Lee, L.-F. (2013), 'Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances', *Regional Science and Urban Economics* **43**, 590–616.
- Jin, F. and Lee, L.-f. (2015), 'On the bootstrap for Moran's I test for spatial dependence', *Journal of Econometrics* **184**, 295–314.
- Kelejian, H. H. (2008), 'A spatial J-test for model specification against a single or a set of non-nested alternatives', *Letters in Spatial and Resource Sciences* **1**, 3–11.
- Kelejian, H. H. and Piras, G. (2011), 'An extension of Kelejian's J-test for non-nested spatial models', *Regional Science and Urban Economics* **41**, 281–292.
- Kelejian, H. H. and Prucha, I. R. (1998), 'A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *Journal of Real Estate Finance and Economics* **17**, 99–121.
- Kelejian, H. H. and Prucha, I. R. (1999), 'A generalized moments estimator for the autoregressive parameter in a spatial model', *International Economic Review* **40**, 509–533.
- Kelejian, H. H. and Prucha, I. R. (2007), 'Relative efficiencies of various predictors in spatial econometric models containing spatial lags', *Regional Science and Urban Economics* **37**, 363–374.
- Kelejian, H. H. and Prucha, I. R. (2010), 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics* **157**, 53–67.
- Kelejian, H. H., Prucha, I. R. and Yuzefovich, Y. (2004), Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results, in J. LeSage and K. R. Pace, eds, 'Spatial and Spatiotemporal Econometrics, Advance in Econometrics', Vol. 18, Emerald Group Publishing, pp. 163–198.

- Lacombe, D. (2004), ‘Does econometric methodology matter? an analysis of public policy using spatial econometric techniques’, *Geographical Analysis* **36**, 105–118.
- Lee, L.-F. (2003), ‘Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances’, *Econometric Reviews* **22**, 307–335.
- Lee, L.-F. (2007), ‘GMM and 2SLS estimation of mixed regressive, spatial autoregressive models’, *Journal of Econometrics* **137**, 489–514.
- Lee, L.-F. and Liu, X. (2010), ‘Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances’, *Econometric Theory* **26**, 187–230.
- Lee, S.-Y. (2008), Three essays on spatial econometrics and empirical industrial organization, PhD thesis.
- LeSage, J. and Pace, R. (2014), ‘The biggest myth in spatial econometrics’, *Econometrics* **2**(4), 217–249.
- Lin, X. and Lee, L.-F. (2010), ‘GMM estimation of spatial autoregressive models with unknown heteroskedasticity’, *Journal of Econometrics* **157**, 34–52.
- Liu, R. (1988), ‘Bootstrap procedures under some non-i.i.d, models’, *Annals of Statistics* **16**(4), 1696–1708.
- Liu, S. F. and Yang, Z. (2015), ‘Modified qml estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality’, *Regional Science and Urban Economics* **52**, 50–70.
- Liu, X., Lee, L.-F. and Bollinger, C. R. (2010), ‘An efficient GMM estimator of spatial autoregressive models’, *Journal of Econometrics* **159**, 302–319.
- Liu, X., Patacchini, E. and Zenou, Y. (2014), ‘Endogenous peer effects: local aggregate of global average?’, *Journal of Economic Behavior and Organization* **103**, 39–59.
- MacKinnon, J. D., White, H. and Davidson, R. (1983), ‘Tests for model specification in the presence of alternative hypotheses’, *Journal of Econometrics* **21**, 53–70.
- MacKinnon, J. G. (2002), ‘Bootstrap inference in econometrics’, *Canadian Journal of Economics* **35**, 615–645.
- Mammen, E. (1993), ‘Bootstrap and wild bootstrap for high dimensional linear models’, *Annals of Statistics* **21**, 255–285.

- Mankiw, G. N., Romer, D. and Weil, D. N. (1992), 'A contribution to the empirics of economic growth', *The Quarterly Journal of Economics* **107**, 407–437.
- Melitz, J. and Toubal, f. (2014), 'Native language, spoken language, translation and trade', *Journal of International Economics* **93**, 351–363.
- Nei, M. (1973), 'Analysis of gene diversity in subdivided populations', *Proceedings of the National Academy of Sciences of the United States of America* **70**, 3321–3323.
- Pace, K. R. and LeSage, J. (2002), 'Semiparametric maximum likelihood estimates of spatial dependence', *Geographical Analysis* **34**(1), 76–90.
- Pesaran, H. M. (1974), 'On the general problem of model selection', *The Review of Economic Studies* **41**, 153–171.
- Pesaran, H. M. and Deaton, A. (1978), 'Testing non-nested nonlinear regression models', *Econometrica* **46**, 677–694.
- Pesaran, H. M. and Weeks, M. (2001), Non-nested hypothesis testing: an overview, in B. H. Baltagi, ed., 'A Companion to Theoretical Econometrics', Blackwell Publishers Ltd.
- Putterman, L. and Weil, D. N. (2010), 'Post-1500 population flows and the long-run determinants of economic growth and inequality', *The Quarterly Journal of Economics* **125**, 1627–1682.
- Reynolds, J., Weir, B. S. and C.C., C. (1983), 'Estimation of the coancestry coefficient: Basis for a short-term genetic distance', *Genetics* **105**(3), 776–779.
- Sherell, N. (1990), The estimation and specification of spatial econometric models, PhD thesis, University of Bristol.
- Spolaore, E. and Wacziarg, R. (2009), 'The diffusion of development', *The Quarterly Journal of Economics* **124**, 469–529.
- Spolaore, E. and Wacziarg, R. (2013), 'How deep are the roots of economic development?', *Journal of Economic Literature* **51**, 325–369.
- Wright, S. (1965), 'The interpretation of population structure by f-statistics with special regard to systems of mating', *Evolution* **19**(3), 395–420.

Yang, Z. (2015), ‘Lm tests of spatial dependence based on bootstrap critical values’, *Journal of Econometrics* **185**(1), 33–59.

Table 9: Matrix of significant elasticities at 5% for OECD countries using trade flows (*W3*)

	AUS	AUT	BEL	CAN	DNK	FIN	FRA	GER	GRC	IRL	ITA	JPN
AUS	<b>0.1654</b>	0.0002	0.0004	-	0.0002	0.0003	0.0011	0.0027	0.0000	0.0003	0.0012	0.0064
AUT	-	<b>0.1655</b>	0.0009	-	0.0002	0.0003	0.0016	0.0117	0.0001	-	0.0025	0.0011
BEL	-	0.0003	<b>0.1656</b>	-	0.0002	0.0002	0.0038	0.0054	0.0000	0.0005	0.0013	0.0010
CAN	0.0002	0.0001	-	<b>0.1659</b>	-	0.0001	0.0008	-	0.0000	-	0.0006	0.0028
DNK	-	0.0004	0.0012	-	<b>0.1654</b>	0.0009	0.0020	0.0067	0.0001	0.0004	0.0015	0.0011
FIN	0.0001	0.0004	0.0010	-	0.0009	<b>0.1654</b>	0.0016	0.0051	0.0001	0.0003	0.0014	0.0020
FRA	-	0.0003	0.0023	-	0.0002	0.0002	<b>0.1657</b>	0.0054	0.0001	0.0004	0.0028	0.0012
GER	-	0.0015	0.0024	-	0.0007	0.0004	0.0039	<b>0.1662</b>	0.0002	0.0006	0.0032	0.0023
GRC	-	0.0003	0.0009	-	0.0003	0.0002	0.0020	0.0039	<b>0.1653</b>	0.0002	0.0036	0.0011
IRL	-	-	0.0005	-	0.0002	0.0002	0.0013	0.0020	0.0000	<b>0.1654</b>	0.0007	0.0016
ITA	0.0001	0.0006	0.0013	-	0.0002	0.0002	0.0032	0.0051	0.0001	0.0003	<b>0.1656</b>	0.0008
JPN	0.0008	0.0002	-	0.0011	0.0003	0.0001	0.0010	0.0022	-	0.0003	0.0009	<b>0.1660</b>
NLD	-	0.0003	0.0031	-	0.0003	0.0003	0.0022	0.0063	0.0000	0.0004	0.0012	0.0016
NZL	0.0058	0.0001	0.0003	-	0.0002	0.0002	0.0008	0.0017	-	0.0002	0.0009	0.0049
NOR	0.0001	0.0004	0.0009	0.0008	0.0021	0.0011	0.0015	0.0046	0.0001	0.0005	0.0013	0.0019
PRT	-	0.0001	0.0006	-	0.0001	0.0001	0.0017	0.0024	0.0000	0.0001	0.0014	0.0005
ESP	-	0.0003	0.0009	-	0.0002	0.0002	0.0037	0.0037	0.0001	0.0003	0.0022	0.0009
SWE	-	0.0005	0.0013	-	0.0019	0.0018	0.0021	0.0062	0.0001	0.0004	0.0013	0.0014
CHE	0.0001	0.0014	0.0013	-	0.0004	0.0003	0.0036	0.0105	0.0001	0.0005	0.0033	0.0014
GBR	0.0002	0.0003	0.0016	0.0006	0.0005	0.0005	0.0031	0.0050	0.0001	0.0013	0.0018	0.0021
USA	0.0002	0.0001	0.0004	0.0054	0.0001	0.0001	0.0011	0.0023	0.0000	0.0003	0.0010	0.0063
ARG	0.0001	0.0001	0.0003	-	0.0001	0.0001	0.0010	0.0013	0.0000	-	0.0011	0.0009
BOL	-	-	0.0001	-	0.0001	0.0001	-	0.0007	-	-	0.0004	0.0013
BRA	0.0000	0.0001	0.0003	0.0004	0.0001	0.0001	0.0006	0.0017	0.0000	0.0001	0.0009	0.0011
BFA	0.0000	-	0.0000	-	0.0000	-	0.0002	-	0.0000	0.0000	0.0000	0.0000
CHL	0.0001	-	0.0003	-	0.0001	0.0002	0.0010	0.0015	-	0.0001	0.0009	0.0017
CHN	0.0001	0.0001	0.0001	-	-	0.0001	0.0005	0.0011	-	-	0.0005	0.0038
COL	-	-	0.0001	0.0004	0.0000	0.0001	0.0004	0.0009	-	-	0.0004	0.0010
CRI	-	-	0.0001	-	0.0000	-	0.0004	0.0005	-	-	0.0003	0.0010
CUB	-	-	0.0003	0.0007	0.0001	-	0.0007	0.0006	-	-	0.0009	-
CYP	-	0.0002	0.0004	-	0.0002	0.0001	0.0010	0.0024	0.0013	0.0002	0.0017	0.0018
ECU	-	-	0.0002	-	-	-	0.0006	0.0010	-	-	0.0007	0.0015
EGY	0.0001	0.0001	0.0003	-	0.0001	0.0001	0.0010	0.0013	0.0001	0.0001	0.0012	0.0007
SLV	-	-	-	-	0.0000	0.0001	0.0004	0.0005	-	-	0.0002	0.0007
HND	-	-	0.0001	-	-	-	-	-	0.0000	-	-	0.0007
HUN	-	0.0027	0.0009	-	0.0002	0.0003	0.0014	0.0078	0.0001	0.0002	0.0023	0.0011
IND	0.0001	0.0001	0.0002	0.0002	0.0001	0.0001	0.0004	0.0010	0.0000	-	0.0004	0.0009
JAM	0.0001	-	0.0002	-	-	0.0001	0.0004	-	-	0.0001	-	0.0015
MDG	0.0000	-	0.0002	-	0.0000	-	0.0019	0.0004	-	-	0.0002	0.0004
MYS	0.0005	-	-	-	-	0.0001	0.0005	0.0009	-	0.0002	0.0003	0.0049
MUS	0.0004	-	0.0003	-	0.0001	-	0.0025	0.0010	0.0000	0.0001	0.0006	0.0012
MEX	-	-	-	-	-	-	-	0.0007	-	-	-	-
NIC	-	-	0.0001	-	0.0001	0.0001	0.0002	0.0003	-	-	0.0002	0.0009
PAN	-	-	-	-	0.0001	-	0.0003	-	0.0001	-	-	0.0096
PRY	-	-	-	-	-	-	0.0003	0.0005	-	0.0001	0.0003	0.0010
PER	-	0.0001	0.0002	-	0.0001	0.0001	0.0004	0.0009	-	-	0.0006	0.0015
PHL	0.0005	-	0.0002	-	-	0.0001	0.0005	0.0010	-	0.0002	0.0003	0.0053
ROM	-	0.0008	0.0006	-	0.0002	0.0001	0.0018	0.0049	0.0004	-	0.0039	-
SGP	0.0003	-	-	-	0.0001	0.0001	0.0006	0.0009	-	0.0001	0.0004	0.0042
ZAF	0.0003	0.0001	0.0004	-	0.0001	0.0001	0.0008	0.0024	-	0.0002	0.0007	0.0013
SYR	-	0.0002	0.0005	-	0.0001	0.0001	0.0010	0.0015	0.0001	-	0.0014	0.0009
THA	0.0003	0.0001	0.0002	-	0.0001	0.0001	0.0005	0.0010	-	-	0.0003	0.0051
TTO	0.0001	0.0001	-	0.0012	0.0001	-	-	0.0010	-	0.0002	0.0005	0.0015
TUN	-	-	0.0004	-	0.0000	0.0000	0.0019	0.0010	0.0000	-	0.0013	0.0002
TUR	0.0000	0.0002	0.0004	-	0.0001	0.0001	0.0011	0.0024	0.0001	0.0001	0.0014	0.0007
UGA	0.0000	0.0001	0.0002	0.0001	0.0001	0.0000	0.0003	0.0004	0.0000	0.0000	0.0003	0.0006
URY	0.0000	-	-	-	0.0001	0.0001	0.0008	0.0009	0.0000	-	0.0010	0.0007
VEN	-	-	0.0002	-	0.0001	0.0001	0.0004	0.0008	-	-	0.0007	0.0008

- denotes non significant (at the 5% threshold) while all figures are significant impact at the 5% level.

Table 9: Matrix of significant elasticities for OECD countries using trade flows (W3), Cont'd

	NLD	NZL	NOR	PRT	ESP	SWE	CHE	GBR	USA	BRA	CHN
AUS	0.0005	0.0012	0.0001	0.0001	0.0003	0.0007	0.0005	0.0023	0.0063	0.0002	0.0019
AUT	0.0010	-	-	0.0002	0.0004	0.0005	0.0011	0.0010	0.0015	-	0.0004
BEL	0.0040	0.0000	0.0001	0.0002	0.0005	0.0007	0.0004	0.0019	0.0018	0.0002	0.0004
CAN	-	0.0001	0.0001	-	-	0.0003	0.0002	0.0009	0.0180	0.0002	-
DNK	0.0021	0.0000	0.0009	0.0003	0.0005	0.0034	0.0006	0.0024	0.0016	0.0001	0.0007
FIN	0.0012	-	0.0007	0.0002	0.0004	0.0033	0.0006	0.0022	0.0023	0.0001	0.0006
FRA	0.0014	0.0000	0.0001	0.0003	0.0016	0.0005	0.0007	0.0021	0.0024	0.0002	0.0006
GER	0.0028	0.0000	0.0003	0.0004	0.0010	0.0008	0.0015	0.0024	0.0028	0.0002	0.0010
GRC	0.0015	0.0000	0.0001	0.0001	0.0008	0.0004	0.0004	0.0015	0.0010	0.0001	0.0004
IRL	0.0009	-	0.0001	0.0001	0.0004	0.0004	0.0002	0.0080	0.0037	-	-
ITA	0.0015	0.0000	0.0001	0.0001	0.0010	0.0004	0.0009	0.0016	0.0013	0.0002	0.0005
JPN	-	0.0002	0.0001	-	0.0002	0.0004	0.0005	0.0011	0.0085	0.0003	0.0044
NLD	<b>0.1656</b>	-	0.0002	0.0002	0.0006	0.0008	0.0005	0.0026	0.0026	0.0003	0.0006
NZL	0.0004	<b>0.1654</b>	0.0001	-	0.0002	0.0006	0.0004	0.0018	0.0045	0.0001	0.0013
NOR	0.0015	-	<b>0.1654</b>	0.0003	0.0005	0.0046	0.0005	0.0029	0.0024	0.0001	0.0007
PRT	0.0008	0.0000	0.0001	<b>0.1654</b>	0.0029	0.0002	0.0002	0.0010	-	0.0002	-
ESP	0.0009	-	0.0001	0.0006	<b>0.1654</b>	0.0004	0.0003	0.0016	0.0014	0.0001	0.0005
SWE	0.0020	-	0.0015	0.0003	0.0005	<b>0.1655</b>	0.0006	0.0030	0.0024	0.0001	0.0005
CHE	0.0016	-	0.0001	0.0002	0.0005	0.0006	<b>0.1655</b>	0.0016	0.0022	0.0001	0.0005
GBR	0.0023	0.0001	0.0003	0.0003	0.0009	0.0009	0.0005	<b>0.1657</b>	0.0038	0.0002	0.0006
USA	0.0004	0.0001	0.0001	0.0001	0.0003	0.0004	0.0004	0.0014	<b>0.1666</b>	0.0005	0.0017
ARG	0.0003	0.0000	-	0.0000	0.0008	0.0002	0.0002	0.0005	0.0038	0.0038	0.0006
BOL	-	0.0000	0.0000	-	0.0003	0.0003	0.0001	-	0.0029	0.0034	-
BRA	0.0003	0.0000	-	0.0001	0.0003	0.0003	0.0003	0.0005	0.0040	<b>0.1654</b>	0.0003
BFA	0.0000	-	-	-	0.0000	-	-	0.0000	0.0000	0.0000	0.0000
CHL	-	0.0001	0.0001	0.0001	0.0008	0.0004	0.0003	0.0005	0.0052	0.0019	0.0009
CHN	0.0002	0.0000	0.0000	-	0.0001	0.0002	0.0001	0.0003	0.0013	0.0001	<b>0.1655</b>
COL	0.0002	-	0.0000	-	0.0003	0.0002	0.0002	0.0004	0.0044	0.0005	-
CRI	0.0002	-	0.0000	-	0.0002	-	0.0001	-	0.0063	0.0004	-
CUB	0.0003	0.0001	-	-	0.0025	0.0001	0.0001	0.0003	-	0.0003	0.0025
CYP	0.0005	0.0000	0.0003	0.0001	0.0006	0.0003	0.0003	0.0020	0.0019	0.0001	0.0004
ECU	0.0002	-	-	-	0.0005	0.0001	0.0002	0.0003	0.0048	0.0008	-
EGY	0.0003	0.0000	0.0000	-	0.0003	0.0003	0.0003	0.0006	0.0013	0.0002	0.0002
SLV	0.0001	0.0000	-	-	0.0002	0.0001	0.0001	-	0.0038	0.0002	-
HND	0.0002	0.0000	-	-	0.0002	-	0.0001	-	0.0055	0.0002	-
HUN	0.0009	-	-	0.0001	0.0004	0.0005	0.0007	0.0010	0.0012	0.0002	0.0006
IND	0.0002	0.0000	0.0000	-	0.0001	0.0001	0.0003	0.0006	0.0011	0.0001	0.0003
JAM	0.0002	0.0001	-	-	0.0001	0.0001	-	0.0009	0.0089	0.0004	-
MDG	0.0001	0.0000	-	-	0.0001	0.0000	0.0001	0.0002	0.0003	0.0001	0.0003
MYS	0.0002	0.0001	-	-	-	0.0002	0.0002	0.0007	0.0024	0.0001	0.0006
MUS	-	0.0002	-	0.0000	0.0002	-	0.0003	0.0011	-	0.0001	0.0009
MEX	-	0.0000	-	-	0.0002	-	0.0001	-	0.0110	0.0002	-
NIC	0.0002	0.0000	0.0000	-	0.0003	0.0001	0.0001	-	0.0032	0.0001	-
PAN	-	-	0.0002	-	0.0003	-	0.0002	-	0.0026	0.0002	-
PRY	-	-	-	-	0.0002	0.0001	0.0001	0.0004	0.0023	0.0042	-
PER	0.0002	0.0002	0.0000	-	0.0005	0.0003	0.0002	0.0003	0.0047	0.0010	0.0006
PHL	0.0003	0.0001	-	-	0.0001	0.0001	0.0002	0.0005	0.0031	0.0002	0.0007
ROM	0.0008	-	-	-	0.0003	0.0003	0.0005	0.0010	0.0010	0.0002	0.0004
SGP	0.0002	0.0000	0.0000	-	0.0001	0.0001	0.0002	0.0006	0.0025	0.0001	0.0006
ZAF	0.0004	0.0000	-	0.0001	0.0002	0.0002	0.0004	0.0016	0.0017	0.0002	0.0005
SYR	0.0004	0.0000	-	-	0.0003	0.0002	0.0002	0.0005	0.0006	0.0001	0.0004
THA	0.0002	0.0001	0.0000	-	0.0001	0.0002	0.0002	0.0004	0.0022	0.0001	0.0007
TTO	0.0004	0.0001	0.0001	-	0.0002	-	0.0002	0.0017	0.0096	0.0004	-
TUN	0.0002	0.0000	-	0.0000	0.0003	0.0001	0.0001	0.0002	0.0003	0.0000	-
TUR	0.0005	0.0000	0.0001	-	0.0004	0.0003	0.0004	0.0008	0.0011	0.0001	0.0003
UGA	0.0002	-	0.0000	-	0.0001	0.0001	0.0001	0.0009	0.0004	0.0000	0.0002
URY	-	0.0000	-	0.0000	0.0007	0.0002	0.0002	0.0004	0.0022	0.0042	0.0004
VEN	0.0002	0.0001	0.0000	-	0.0003	0.0001	0.0001	0.0004	0.0052	0.0005	-

- denotes non significant (at the 5% threshold) while all figures are significant impacts at the 5% level.

## Appendix A. RGMM estimation of heteroskedastic SAR models

The objective of this Appendix is to present the GMM estimation of a SAR specification in the presence of unknown heteroskedasticity, developed by Lin and Lee (2010). The model is recalled in equation (A.1).

$$Y_n = X_n\beta_0 + \lambda_0 W_n Y_n + \varepsilon_n. \quad (\text{A.1})$$

We also  $\theta_0 = [\lambda_0, \beta_0']'$ ,  $\theta = [\lambda, \beta']'$  for any value of  $\lambda$  and  $\beta$ ,  $S_n = (I_n - \lambda_0 W_n)$ ,  $S_n(\lambda) = (I_n - \lambda W_n)$  and we define  $\varepsilon_n(\theta) = S_n(\lambda)Y_n - X_n\beta$  for any value of  $\theta$ . Finally, we define  $A_n^s = A_n + A_n'$  for any square matrix  $A_n$ .

Assuming the existence of  $S_n^{-1}$ , the reduced form of model (A.1) is as follows

$$Y_n = S_n^{-1} X_n \beta + S_n^{-1} \varepsilon_n, \quad (\text{A.2})$$

From (A.2), we can express  $W_n Y_n$  as follows:

$$W_n Y_n = G_n X_n \beta_0 + G_n \varepsilon_n, \quad (\text{A.3})$$

with  $G_n = W_n S_n^{-1}$ .

In general,  $W_n Y_n$  is correlated with  $\varepsilon_n$  since  $\mathbf{E}((W_n Y_n)' \varepsilon_n) = \text{tr}(\Sigma_n G_n) \neq 0$  where  $\Sigma_n$  is a diagonal matrix containing the error term variance  $\sigma_{in}^2 = \mathbf{E}(\varepsilon_{in}^2)$  for  $i = 1, \dots, n$ .

The RGMM approach developed by Lin and Lee (2010) combines linear and quadratic moments. Let  $Q_n$  be a  $n \times k^*$  matrix of IV's, with  $k^* \geq k + 1$ , which contains the information of the deterministic part of  $W_n Y_n$  in (A.3). This matrix is constructed from  $X_n$  and functions of  $W_n$  and is used to instrument  $W_n Y_n$ . For instance,  $Q_n = [X_n, W_n X_n, W_n^2 X_n, \dots]_{LI}$ , where  $LI$  stands for linearly independent columns.

Lin and Lee (2010) additionally consider quadratic moments  $(P_{jn} \varepsilon_n(\theta))' \varepsilon_n(\theta)$ ,  $j = 1, \dots, m$ , that capture the information contained in the stochastic part of (A.3). These quadratic moments serve as IV's for  $G_n \varepsilon_n$ , the stochastic part of  $W_n Y_n$ . One needs to impose  $P_{jn} \varepsilon_n$  to be uncorrelated with  $\varepsilon_n$  for the quadratic moments to act as orthogonality conditions. Under heteroskedasticity of the error term,  $P_{jn}$

should belong to the class  $\mathcal{P}_{2n}$  that contains the constant  $n \times n$  matrices with zero diagonal.<sup>27</sup>

Quadratic moments are used for two reasons. As they capture the information contained in the stochastic part of (A.3), they make the obtained estimator more efficient compared to TSLS or BTSLS estimators. Second, if all  $X_n$  are irrelevant, linear moments can not be used to instrument  $W_n Y_n$ . However, in that case, which corresponds to the pure SAR model, i.e.  $Y_n = \lambda W_n Y_n + \varepsilon_n$ , quadratic moments can still act as instruments for  $W_n Y_n$ .

Assuming  $m$  quadratic moments, the set of moment functions is

$$\begin{aligned} g_n(\theta) &= (P_{1n}\varepsilon_n(\theta), \dots, P_{mn}\varepsilon_n(\theta), Q_n)' \varepsilon_n(\theta) \\ &= (\varepsilon_n(\theta)' P_{1n} \varepsilon_n(\theta), \dots, \varepsilon_n(\theta)' P_{mn} \varepsilon_n(\theta), \varepsilon_n(\theta)' Q_n)' \end{aligned} \quad (\text{A.4})$$

By contrast to maximum likelihood estimation,  $\lambda$  is not required to belong to some compact parameter space.<sup>28</sup> In the QML framework, the constraints imposed on the parameter space ensure the positiveness of the Jacobian of the transformation  $|I_n - \lambda W_n|$ . However, in the GMM framework, such constraints are not relevant because the objective function is a polynomial function (of order 4) in  $\theta$ . The condition on the boundedness of the parameter space is solely for the theoretical purpose of proving consistency of the GMME (Lin and Lee, 2010). Hence, the GMM is a unconstrained estimation procedure.

Instead of minimizing the objective function  $g_n(\theta)' g_n(\theta)$  with respect to  $\theta$ , Lin and Lee (2010) propose to minimize a linear combination  $a_n g_n(\theta)$  of the moment functions  $g_n(\theta)$ , so that the objective function enters the Hansen (1982) setting where each condition is weighted by its precision. The matrix  $a_n$  has a full rank greater than or equal to the dimension of  $\theta$ . The objective function is thus  $g_n(\theta)' a_n' a_n g_n(\theta)$  where the matrix  $a_n' a_n$  is non-negative and contains the weight of each orthogonality condition and motivates the issue of optimal weighting.

Under some Assumptions (not reported here), Lin and Lee (2010) have shown that the Robust GMME

---

<sup>27</sup> If we impose homoskedasticity, the matrices  $P_{jn}$  may belong to the wider class  $\mathcal{P}_{1n}$  that contains the constant  $n \times n$  matrices with zero trace. Depending on the assumptions made on the error term, matrices  $P_{jn}$  may belong to  $\mathcal{P}_{1n}$  or  $\mathcal{P}_{2n}$ .

<sup>28</sup> For more information about the parameter space for  $\lambda$ , the interested reader may consult Kelejian and Prucha (2010).



$\hat{\theta}$  is a consistent estimator of  $\theta_0$  and  $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, \Gamma)$ , where

$$\Gamma = \lim_{n \rightarrow \infty} \left( \frac{1}{n} D_n' a_n' a_n D_n \right)^{-1} D_n' a_n' a_n \Omega_n a_n' a_n D_n (D_n' a_n' a_n D_n)^{-1} \quad (\text{A.5})$$

with

$$D_n = \frac{\partial \mathbf{E}(g_n(\theta_0))}{\partial \theta'} = - \begin{pmatrix} \text{tr}(\Sigma_n P_{1n}^s G_n) & 0 \\ \vdots & \vdots \\ \text{tr}(\Sigma_n P_{mn}^s G_n) & 0 \\ Q_n' G_n X_n \beta_0 & Q_n' X_n \end{pmatrix} \quad (\text{A.6})$$

$$\Omega_n = \text{Var}(g_n(\theta_0)) = \begin{pmatrix} \text{tr}(\Sigma_n P_{n,1}(\Sigma_n P_{1n})^s) & \text{tr}(\Sigma_n P_{1n}(\Sigma_n P_{2n})^s) & \cdots & 0 \\ \text{tr}(\Sigma_n P_{n,2}(\Sigma_n P_{1n})^s) & \text{tr}(\Sigma_n P_{2n}(\Sigma_n P_{2n})^s) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & Q_n' \Sigma_n Q_n \end{pmatrix} \quad (\text{A.7})$$

Even though the choice of the weighting matrix  $a_n' a_n$  does not affect the consistency of RGMME, it plays a role in its efficiency. By setting  $a_n' a_n$  equal to the inverse of the covariance matrix of orthogonality conditions ( $\Omega_n$ ), one obtains the optimal RGMME, i.e. the most efficient estimator given  $g_n(\theta)$ . In their proposition 3, Lin and Lee (2010) show that if one has a consistent estimator  $\hat{\Omega}_n$ , the asymptotic distribution of the feasible “optimal” RGMME obtained from minimizing  $g_n(\theta)' \hat{\Omega}_n g_n(\theta)$  with respect to  $\theta$  is as follows:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N \left( 0, \left( \lim_{n \rightarrow \infty} \frac{1}{n} D_n' \Omega_n^{-1} D_n \right)^{-1} \right) \quad (\text{A.8})$$

The optimal estimator is the most efficient estimator that can be obtained given a set of orthogonality conditions. However, we could question the existence of the best choice of orthogonality conditions, i.e. the set of conditions which guarantees that the best optimal RGMME is the most efficient one. Unfortunately, in the unknown heteroskedasticity case, no such set of orthogonality conditions exists, since they would involve the unknown  $\Sigma_n$ . However, Lin and Lee (2010) suggest to use as orthogonality conditions those corresponding to the best conditions in the *iid* case when  $P_{jn} \in \mathcal{P}_{n,2}$ . These conditions, shown in Lee (2007), are  $P_n = (G_n - \text{Diag}(G_n))$  and  $Q_n = [G_n X_n \beta_0, X_n]$ , where  $\text{Diag}(A_n)$  refers to a diagonal matrix

where diagonal elements correspond to the diagonal of  $A_n$ .<sup>29</sup> Hence, they suggest to use only one quadratic and  $k + 1$  linear conditions. The vector of orthogonality conditions becomes:

$$g_n(\theta) = (\varepsilon'_n(\theta)P_n\varepsilon_n(\theta), \varepsilon'_nQ_n)' \quad (\text{A.9})$$

Expressions for the gradient and the covariance matrix of  $g_n(\theta)$  defined in (A.9), evaluated at the true parameter value are:

$$D_n = -\frac{\partial \mathbf{E}(g_n(\theta_0))}{\partial(\theta)} = \begin{pmatrix} \text{tr}(P_n^s G_n \Sigma_n) & 0 \\ Q'_n G_n X_n \beta_0 & Q'_n X_n \end{pmatrix} \quad (\text{A.10})$$

$$\Omega_n = \text{Var}(g_n(\theta_0)) = \begin{pmatrix} \text{tr}(\Sigma_n P_n (\Sigma_n P_n)^s) & 0 \\ 0 & Q'_n \Sigma_n Q_n \end{pmatrix} \quad (\text{A.11})$$

As these orthogonality conditions depend on the unknown parameters  $\lambda_0$  and  $\beta_0$ , we need to iterate to get the optimal RGMME. The first step consists in getting consistent estimators of  $\lambda_0$  and  $\beta_0$ . To do so, we compute RGMME based on orthogonality conditions not involving any unknown parameters but which satisfy the above assumptions. For instance, we set  $P_n = W_n$  and  $Q_n = [X_n, W_n X_n, W_n^2 X_n]$ . Also we use as initial guess of the weight matrix  $a'_n a_n$  the identity matrix. The obtained (first-step) estimators,  $\check{\lambda}^{fs}$  and  $\check{\beta}^{fs}$ , are then used to compute the residuals which allow to have an estimator  $\check{\Sigma}_n$ , whose  $i^{th}$  diagonal element is  $\check{\varepsilon}_{in}^2(\check{\theta}^{fs})$ . We are now able to compute an estimator  $\check{\Omega}_n$  of  $\Omega_n$ . Besides, we can construct the new orthogonality conditions  $g_n^{st} = (\varepsilon'_n(\theta)P_n^{st}\varepsilon_n(\theta), \varepsilon'_n(\theta)Q_n^{st})'$  based on  $P_n^{st} = (\check{G}_n - \text{Diag}(\check{G}_n))$  and  $Q_n^{st} = [\check{G}_n X_n \check{\beta}^{fs}, X_n]$ , where  $\check{G}_n = W_n(I_n - \check{\lambda}^{fs} W_n)^{-1}$ .

The next step consists in iterating between estimation of  $\lambda_0$  and  $\beta_0$  on the one hand and computation of an estimator of  $\Omega_n$  on the other hand until we reach convergence.<sup>30</sup>

Statistical inference will be based on the estimator  $\hat{\theta}$  obtained after convergence has been reached, which has an estimated asymptotic covariance matrix equal to  $\left(\frac{1}{n}\hat{D}'_n \hat{\Omega}_n^{-1} \hat{D}_n\right)^{-1}$ . In this expression,  $\hat{D}_n$  is constructed from plugging the final-step estimators within equation (A.10) and  $\hat{\Omega}_n$  comes from plugging

<sup>29</sup> If quadratic moments  $P_n$  belong to  $\mathcal{P}_{n1}$ , Liu et al. (2010) derive the best moment conditions under homoskedasticity, no matter the distribution of the error terms.

<sup>30</sup> In this paper, the following convergence criteria has been chosen: we iterate until  $\sum_{j=1}^{K+1} |\hat{\theta}_{j,s+1} - \hat{\theta}_{j,s}| < 0.0001$ , where  $\hat{\theta}_{j,s}$  is the estimator of  $\theta_{0,j}$  obtained after  $s$  iterations.

the final-step estimators in equation (A.11).